

Reverse Bayesianism: Revising Beliefs in Light of Unforeseen Events

Christoph K. Becker¹, Tigran Melkonyan², Eugenio Proto³, Andis Sofianos¹, and Stefan T. Trautmann¹

¹*University of Heidelberg*

²*University of Alabama, Tuscaloosa*

³*University of Glasgow, CEPR, IZA and CesIfo*

May 7, 2022

Abstract

Bayesian updating is the dominant theory of learning. However, the theory is silent about how individuals react to events that were previously unforeseeable or unforeseen. We test if subjects update their beliefs according to “*reverse Bayesianism*”, under which the relative likelihoods of prior beliefs remain unchanged after an unforeseen event materializes. Across two experiments we find that participants do not systematically deviate from reverse Bayesianism. However, we do find well-known violations of Bayesian updating. Furthermore, decision makers vary in their ex-ante unawareness depending on the context.

JEL classification: C11, C91, D83, D84

Keywords: Reverse Bayesianism, Unforeseen, Unawareness, Bayesian Updating

Acknowledgements

The authors thank several colleagues for valuable input. In particular, Ala Avoyan, Andrew Caplin, Stefano Caria, David Cesarini, Cary Deck, Jürgen Eichberger, Guillaume Frechette, Peter Hammond, Andrea Isoni, Edi Karni, Christian König-Kersting, Fabian Paetzel, Joerg Oechssler, Daniela Puzzello, Andy Schotter, Daniel Sgroi, and Stefan Traub provided numerous insights. We also thank seminar and conference participants at the Indiana University, New York University, University of Sussex, the ESA 2020 Global, CESifo Area Conference on Behavioral Economics 2020, Annual Canadian Economics Association Meeting 2021, RUD 2021, and Experimental Finance 2021 for helpful comments. We are also very grateful to J. Philipp Reiss for the use of the Karlsruhe Decision & Design Lab. Both experiments were pre-registered in the AEA Registry. The University of Glasgow, University of Heidelberg, and University of Warwick provided funding for this research.

Corresponding Author: Eugenio Proto, eugenio.proto@gmail.com

1 Introduction

We live in a world where scientific progress, human activities, and events outside of our control constantly lead to discoveries and observations of *unforeseen*¹ and *unforeseeable*² phenomena that fundamentally change our worldviews and behavior. Situations with unforeseen or unforeseeable events are abundant.³ Even when we can imagine rough outlines of a phenomenon or even when we have a rather precise understanding of its characteristics, we often overlook it in the construction of our universe or include it in the description of the universe but render it as impossible. Examples of such phenomena include global pandemics, political and economic crises, as well as scientific groundbreaking discoveries.

In some cases, the distinction between unforeseeable and unforeseen is blurred and individual-specific. As a result of differences in knowledge and cognitive capacity, what is foreseeable for some people might be unforeseeable to others. A very notable example of an event that was foreseeable but unforeseen by many is COVID-19. Numerous scientists and observers, including Bill Gates, have repeatedly warned us about the possibility of a disastrous pandemic. However, these warnings have been largely ignored by many policy makers, public health officials, and economic decision-makers. The financial crisis of 2007 had a similar nature. For example, Lehman Brothers assumed in 2005 that the worst-case scenario in the housing market was a temporary price depreciation of 5% over the next three years, followed by a rebound and price increase of 5% thereafter; a scenario with a substantial drop in prices over an extended period was not even considered (Gennaioli and Shleifer, 2018, p. 52).

Given the empirical relevance of unawareness and neglected events, the current paper aims to provide insight into how people update their beliefs when unforeseen events materialize. In such cases, Bayes rule is silent about how individuals update their beliefs and is not useful in formulating subsequent reactions. A number of different approaches have been advanced to examine behavior under such circumstances. The epistemic and choice-theoretic approaches are the two main strands in the lit-

¹We call an event unforeseen if either (i) a decision-maker is aware that the event may occur but assigns zero probability to it or (ii) she is not aware that the event may occur, but in principle could be (based on available information).

²We call an event unforeseeable when the information available to a decision maker is objectively insufficient to allow her to contemplate the existence of the event.

³Related concepts are Knightian uncertainty and the “unknown unknowns”, a term famously coined by the late former US Secretary of Defense Donald Rumsfeld.

erature. The goal of the epistemic approach (e.g. Dekel et al., 1998; Modica and Rustichini, 1999; Heifetz et al., 2006; Halpern and Rêgo, 2008; Grant and Quiggin, 2013, among others) is to develop logical approaches and definitions of awareness, unawareness, and partial awareness in non-strategic and strategic settings. By its very own nature, this approach is concerned with laying the epistemic foundations of unawareness rather than producing readily testable hypotheses how decision-makers perceive and react to unforeseen events. The choice-theoretic approach (e.g. Kochov, 2010; Ortoleva, 2012; Karni and Vierø, 2013; Schipper, 2013; Grant and Quiggin, 2015; Grant et al., 2017; Chambers and Hayashi, 2018; Dietrich, 2018; Dominiak and Tserenjigmid, 2021; Schipper, 2022, among others) develops representations of preferences in the presence of unawareness from behavioral axioms on individual preferences. A central property in the literature on decision making under growing awareness is *reverse Bayesianism* (Karni and Vierø, 2013, 2015, 2017; Karni et al., 2020), according to which decision-makers react to prior null events by proportionately shifting probability mass to these events from the prior non-null events. That is, a reverse Bayesian’s construction of a new universe maintains consistency with the old structure. The centrality of reverse Bayesianism stems from a number of factors. First, it is normatively appealing in terms of how information is used to form beliefs. Second, many important models of exchangeable random partitions in statistics and combinatorial decision-theory (e.g. Schipper, 2022) as well as behavioral models of unawareness (e.g. Dominiak and Tserenjigmid, 2021; Piermont, 2021) are either consistent with reverse Bayesianism, or have a non-trivial overlap with it. Furthermore, it is often used as a yardstick against which models of updating beliefs under unawareness are compared. Reverse Bayesianism is also intuitively simple and amenable to testing using behavioral data. In light of all these considerations, we focus our exploration of behavior under unawareness on testing reverse Bayesianism.

Reverse Bayesianism imposes a rationality constraint on the process of updating beliefs following null events. Suppose, for example, a decision-maker bets repeatedly on the color of a randomly drawn marble from an urn which she believes to contain 50 black marbles and 50 white marbles. At one point the decision-maker witnesses some number (known or unknown) of red marbles being unexpectedly added to the urn. Under this design, the contents of the original two-color urn are part of the updated three-color urn. Put differently, the “old world” remains a part of the “new world.” As the information about the old world did not change, a rational updating

rule requires that the decision-maker’s posterior beliefs put equal probability weight on white and black marbles, irrespective of the number of red marbles added to the urn. This is exactly the updating process reverse Bayesianism predicts.

This updating is, however, often less trivial than the above example might suggest. With skewed distributions of beliefs or multiple initial outcomes, keeping likelihood estimates proportional to each other becomes challenging. Complying with the demands of reverse Bayesianism might hence be cognitively demanding in many circumstances. Bayesian updating is commonly violated by decision makers in many situations for similar reasons (Tversky and Koehler, 1994; Sonnemann et al., 2013; Benjamin, 2019). In the context of unforeseen events, descriptive validity may be affected by the asymmetric impact of the new information on the evaluation of existing events. We will discuss different mechanisms for such an asymmetric impact in the next section, including asymmetric salience, $1/N$ -bias, and hindsight bias. It follows from these considerations that whether, and if so, how well decision makers adhere to reverse Bayesianism, is not immediately clear.

The present paper develops two experiments to study the formation of beliefs under growing awareness, and more specifically, to determine whether individuals adhere to reverse Bayesianism. We design our experiments so that the new and unexpected environment retains some parts of the old world. In addition, we test how belief formation and updating are moderated by the environment of the decision situation. According to our knowledge, the present paper is the first to experimentally examine belief formation and reactions to unforeseen events. Furthermore, it is the first experimental study of how expectations of the unknown evolve as the universe expands. We find that behavior in both experiments is consistent with reverse Bayesianism, despite the fact that the participants exhibit some commonly observed judgment biases. Based on our findings, reverse Bayesianism seems to be a natural updating rule for decision-makers, being compelling both from a normative and descriptive perspective.

A controlled laboratory experiment is perhaps the only environment where it is possible to perform our empirical exercise. Unforeseeable events are rare, and by definition it is impossible to predict them and set the stage for observing beliefs in a sufficiently accurate way. At the same time, in the controlled environment of an experiment, it is virtually impossible to generate objectively unforeseeable events. Our experimental designs involve events that vary by the degree of foreseeability. To distinguish them from objectively unforeseeable and objectively foreseeable events,

we coin the events in our experiments as *reasonably unforeseeable* and *reasonably foreseeable*. Whether an event belongs to one of these two latter categories depends on the amount of information received by a participant in the experiment. Our empirical analysis reveals that participants generally do not expect the unknown when it is reasonably unforeseeable. In contrast, some expect an unknown event when it is reasonably foreseeable.

In the first experiment, we analyze behavior of participants who face either a reasonably unforeseeable or a reasonably foreseeable event. In the course of the experiment, we elicit beliefs about the content of an urn as well as willingness to sell a gamble that pays according to a prize randomly drawn from the urn. In both treatments of the experiment, the task entails the introduction of a new urn (with new prize(s)), which was hidden from the participants, and the subsequent addition of its content to the original urn. We find strong evidence for reverse Bayesianism in both treatments. Prior to encountering the surprise in the form of a new urn, participants on average estimate that the probability of a yet unobserved prize is zero. This probability estimate also remains zero after witnessing the surprise, except for some of the participants that were forewarned about the possibility of new prizes in their treatment. We further investigate how the nature of the surprise affects beliefs and the valuation of prospects. Two patterns emerge. First, in the treatment with forewarning about new prizes, higher valuations of prospects are always observed. That is, the possibility of the unknown seems to instill hope, rather than fear. Second, valuations increase (decrease) after a positive (negative) surprise, showing that decision makers do incorporate the new information.

In the second experiment, the new events are all reasonably foreseeable. Participants explore a digital urn by sequentially extracting marbles from it with replacement. They receive no information on which colors are possible, so that the reasonably foreseeable event is represented by so far unobserved colors. This setup allows us to study how their beliefs evolve over time. Participants in the second experiment are found to suffer from common Bayesian updating violations. Nevertheless, we again find that beliefs are updated in accordance with reverse Bayesianism. Additionally, we find that participants lower their perceived likelihood of further unknown events as they sample more or observe more unforeseen colors. Despite this, beliefs about potentially unforeseen events are very persistent, and about one third of the participants still expect a yet unobserved color even after 30 draws.

The rest of the paper is organized as follows. In Section 2 we provide a theoretical framework and derive the hypotheses that are tested in both experiments. In Section 3 we present the design and results of Experiment 1. Section 4 is dedicated to Experiment 2. Section 5 provides a general discussion of the results and concludes the paper. In the appendix we include the experimental instructions and some additional analysis for each experiment.

2 Theoretical Background and Hypotheses

Following Karni and Vierø (2017), let A denote a finite, non-empty set of actions and C_0 denote a finite, non-empty set of *feasible consequences*. To illustrate this framework, consider a pharma company appraising which of two research programs to invest in. Both programs are aimed at developing a drug to treat certain medical condition Y . The set A is given by the two research programs, a and b , the pharma company is considering for investment. The set C_0 represents the consequences of its choice in terms of either developing an effective drug to treat Y or being unsuccessful in that endeavor, denoted by U . Thus, we have $C_0 = \{Y, U\}$ in our example. Let also $x = \neg C_0$ denote an abstract residual consequence, which stands for “something other than what the decision-maker can describe” – for example, finding a treatment for some other medical condition that the pharma company could find a treatment for, but which it is not aware of.

The sets $\hat{C}_0 = C_0 \cup \{x\}$ and A together define the *augmented conceivable state space* via $\hat{C}_0^A := \{s : A \rightarrow \hat{C}_0\}$. That is, the *augmented conceivable state space* takes into account the possibility that an action may lead to the “everything else” consequence x . Moreover, the space of *fully describable conceivable states* is defined as $C_0^A := \{s : A \rightarrow C_0\}$, where the mappings’ image is restricted only to describable consequences.

The augmented conceivable state space can be expanded by observing a new consequence $c' \notin C_0$. In our example, the pharma company may subsequently realize that, as a third consequence, either research program may produce a drug that treats some alternative medical condition Z instead of medical condition Y . The set of feasible consequences then expands to $C_1 = C_0 \cup \{c'\}$. In our example, $C_1 = \{Y, Z, U\}$. Furthermore, C_1^A and \hat{C}_1^A can be defined analogously to C_0^A and \hat{C}_0^A , respectively.

Denote π_0 and π_1 as probability measures defined on C_0^A and C_1^A , respectively,

and representing beliefs before and after a new consequence is observed. In addition to standard axioms guaranteeing an expected utility representation, Karni and Vierø (2017) impose an axiom of invariant risk preferences and two awareness consistency axioms. The latter three axioms ensure that preferences for different levels of awareness are consistent with each other. The resulting representation is characterized by expected utility preferences for different levels of awareness and reverse Bayesianism, with the latter requiring that for all $s, s' \in C_0^A$:

$$\frac{\pi_0(s)}{\pi_0(s')} = \frac{\pi_1(s)}{\pi_1(s')}.$$

That is, this model implies that the decision maker will hold the ratio of probability estimates for known outcomes constant after observing an unforeseen outcome. Under classical Bayesian updating, new information shrinks the state space by excluding some outcomes that had previously been assigned a positive prior probability. In contrast, the present model focuses on the reverse situation, where new information can expand the state space, while still making sure that beliefs are updated in accordance with Bayes rule. Hence, the name “reverse Bayesianism”.

Once a new consequence is discovered, a decision-maker will update her beliefs about further possible new outcomes, now captured by $x = \neg C_1$. Observing a new outcome can have either an increasing or decreasing effect (or none) on the decision maker’s awareness about unforeseen events. On the one hand, it is possible that the discovery of new consequences decreases the amount of remaining unforeseen consequences. On the other hand, discovering a new consequence may highlight that there are still unforeseen consequences to uncover. Accordingly, the model allows for both a decrease or increase in the probability assigned to the residual consequence.

Based on the above framework, we will test if the normative reverse Bayesian model matches the actual behavior of participants in incentivized decision-making experiments. In our study, decision makers state their beliefs about the likelihoods of different events (prizes in Experiment 1 and colors of marbles in Experiment 2) and express their willingness to accept (*WTA*) for the prospects in Experiment 1, using standard incentivizing procedures. The descriptive validity of reverse Bayesianism in this context is not trivial, given the large literature on violations of Bayesian updating. In the context of unforeseen events, descriptive validity may be affected by the asymmetric impact of the new information on the evaluation of existing events

(high versus low prizes in Experiment 1; frequent versus less frequent colored marbles in Experiment 2). Different mechanisms for an asymmetric impact are conceivable. First, experience of unforeseen events may lead to asymmetric salience of different, previously observed prizes or colors. Second, $1/N$ bias may asymmetrically affect events considered more or less likely before observing the new outcome (Sonnemann et al., 2013). Third, hindsight bias has been shown to lead to revisions of ex-ante beliefs: this may loosen the connection between beliefs before and after a new outcome is observed, regarding the previously observed outcomes (Hoffrage and Gigerenzer, 2000).

For Experiment 1, the set of fully describable conceivable states is given by different combinations of prizes in an urn. The probability of an unknown prize is elicited implicitly and a participant’s likelihoods of already observed prizes are not restricted to sum to 1. For Experiment 2, the set of fully describable conceivable states is given by different combinations of colored marbles in an urn. The probability of an unknown event is elicited explicitly by asking participants to state their belief about “any other possible color.”

We differentiate between the *original urn* (before a new outcome is observed) and the *updated urn* (after a new outcome is observed). In the following we denote the probability estimates of each participant for a given state i by \hat{p}_i^o and \hat{p}_i^u , for the original and the updated urn, respectively. The residual estimate is denoted by \hat{p}_x^o and \hat{p}_x^u . Our two experiments will test the following main hypothesis:

Hypothesis 1. *Participants update their beliefs according to reverse Bayesianism. That is, for any \hat{p}_i^o, \hat{p}_i^u and any states $i, i' \in C_0^A$:*

$$\frac{\hat{p}_i^o}{\hat{p}_{i'}^o} = \frac{\hat{p}_i^u}{\hat{p}_{i'}^u}.$$

In the analysis of the two experiments, we will refer to the difference between the ratios before and after the update as:

$$\Delta R = \frac{\hat{p}_i^u}{\hat{p}_{i'}^u} - \frac{\hat{p}_i^o}{\hat{p}_{i'}^o}.$$

In some of our experimental treatments, we explicitly rule out the possibility of unforeseen events and inform the participants about this. If a participant trusts that information, then $\{x\}$ should be empty for her and, as a result, $\hat{C}_0 = C_0$. Specifically, we will test the following:

Hypothesis 2. *At point t of the elicitation, the residual estimate:*

(a) $\hat{p}_x^t = 0$ in the treatments where unforeseen events are ruled out so that $\{x\}$ is empty and $\hat{C}_0 = C_0$.

(b) $\hat{p}_x^t > 0$ in the treatments where unforeseen events are not ruled out.

A further novelty of our experiments is that we explicitly study how residual estimates change after the state space expands. As outlined above, the framework is flexible in regard to how a decision maker’s awareness reacts to a new event. Specifically, it allows for decision makers to increase their residual in either direction after observing a new, unforeseen event. Accordingly, we test the following agnostic hypothesis:

Hypothesis 3. *Participants will not adjust their residual belief after encountering a new event:*

$$\hat{p}_x^u - \hat{p}_x^o = 0.$$

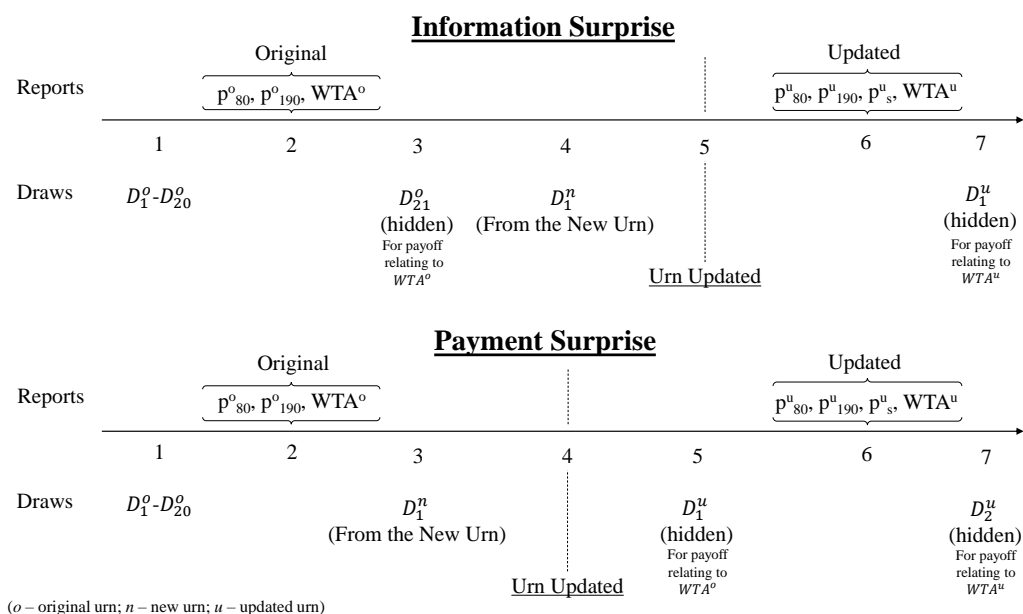
3 Experiment 1

3.1 Design

Experiment 1 elicits beliefs and valuations of prospects before and after encountering a new event. We test reverse Bayesianism using either a reasonably unforeseeable (*Information surprise/IS*) or reasonably foreseeable (*Payment surprise/PS*) event (we omit the word “reasonably” in this section). Each of these two conditions employs either a favorable or adverse new event (*high prize, low prize*), resulting in a 2×2 between-subjects design. Figure 1 provides an overview of the timing in the experiment. We will provide a rationale for our choice of the four conditions after we spell out their details. Our reasoning behind the labels *Information surprise* and *Payment surprise* will also become apparent.

Under the *IS* condition, the participants are presented with an urn, called the *original* urn, and are informed that the urn contains balls with labels representing prizes measured in tokens. Each earned token is exchanged for €0.05 at the end of the experiment. The participants are told that: “*the urn contains two and only two*

Figure 1: The timing of the two surprise conditions



prizes”. However, they are not told what these two prizes or their relative proportions are. Furthermore, we do not alert the participants that the composition of the urn might change by adding or removing balls from the urn.

Following the description of the urn, the participants observe a sequence of 20 physical draws with replacement from the original urn ($D_1^o - D_{20}^o$ in Figure 1). The original urn contains 24 balls labeled ‘80’ and 36 balls labeled ‘190’. No information regarding the specific composition of the urn is disclosed to the participants. All of the draws are made by a participant we refer to as the “experimental assistant,” who is randomly selected for this task from the participants in a given session. The outcome of each draw is revealed to all participants by the experimental assistant.⁴ Thus, everybody in a session observes the same sample. None of these 20 draws are payment-relevant. The only purpose of these draws is for the participants to gain information about the composition of the original urn.

After observing these 20 draws from the original urn, the participants are asked to provide estimates of the probabilities of the prizes they have been observing during these draws (subjective probabilities p_{80}^o and p_{190}^o) and to state their willingness-to-

⁴The experimental assistants do not complete any of the tasks that the other participants perform and receive a fixed payment of €14, which is close to the average earnings in the experiment.

accept (minimum selling price) to sell the prospect of drawing a prize from this urn (WTA^o). We use superscript ‘ o ’ in our notation to emphasize that these values are elicited before any changes to the *original* urn are made. The reported estimates of the two probabilities do not have to add up to 100%. The design does not force the sum of the estimates to be lower or greater than 100% either. We do not, however, explicitly ask for the respondents’ estimates of observing a prize that they have not observed during the 20 draws. Thus, our design allows for an implicit estimation of a residual probability of outcomes that have not yet been observed. This contrasts with Experiment 2 where we explicitly ask for that probability.

Following the 20 draws and the reports of the two probabilities and WTA, a draw from the original urn is made by the experimental assistant (D_{21}^o in Figure 1). The outcome of this draw determines the potential payments for the reports of WTA^o . However, the draw is concealed from the participants when it is made. It is revealed only at the very end of the experiment when the final payment to the participants is displayed, provided that this decision is selected for payment.

Subsequently, we bring out a new urn to the front of the experimental lab. One ball is then drawn from the new urn (D_1^n in Figure 1), revealing a new prize s to all participants. At this point in the experiment, the participants are informed that: “*This urn contains only the prize you are (about to be) shown.*” The value s of the new prize varies with the prize condition. In the *low* prize condition, the new urn contains 15 balls labeled ‘15’, while in the *high* prize condition, it contains 15 balls labeled ‘375’. Although the participants know the value of all prizes in the new urn, they are not informed about how many balls are contained in the new urn. After revealing the value of the prizes in the new urn, we empty its contents into the original urn.⁵ We call this combined urn the *updated* urn. The participants are then asked to estimate the probabilities of each of the three prizes ($p_{80}^u, p_{190}^u, p_s^u$) and to state their WTA for the prospect to draw a prize from the updated urn (WTA^u). We use superscript ‘ u ’ in our notation to emphasize that these values are elicited after the urn is updated. Following these reports, the experimental assistant draws a ball from the updated urn (D_1^u in Figure 1). This draw is concealed and is only revealed at the very end of the experiment, provided this decision is selected for payment.

⁵To prevent the respondents from inferring the number of balls in either urn from the first 20 draws and emptying of the new urn into the original one, all of the balls were made of styrofoam while the boxes were made of opaque plastic material.

Consider now the *PS* condition. Similarly to the *IS* condition, the participants are presented with the *original* urn, and are informed that the urn contains balls with labels that represent prizes. Again, each earned token is exchanged for €0.05 at the end of the experiment. In contrast to *IS*, the participants are not told about the number of different prizes in the *original* urn. Similarly to *IS*, they are not provided with any information about the proportions of balls with any specific prize. In the *PS* condition the respondents are informed that: “*at any point in the study new balls representing different tokens to what you have been observing so far may be added to this urn*”. Thus, one might expect that some participants may incorporate this piece of information, which is not provided under *IS*, into their process of arriving at and reporting probabilities and WTAs. Similarly to *IS*, the participants subsequently observe 20 physical draws from the original urn ($D_1^o - D_{20}^o$ in Figure 1).

After observing the 20 draws from the *original* urn, the participants are asked to report their estimates of the probabilities of the prizes that they have been observing (subjective probabilities p_{80}^o and p_{190}^o) and to state their willingness-to-accept to sell the prospect of drawing a prize from the urn (WTA^o). As for *IS*, the reported estimates of the probabilities are not restricted to add up to 100% or to be smaller or larger than 100%, allowing for calculation of an implicit residual probability.

Following the elicitation of these probabilities and WTA^o , we bring out a new urn to the front of the experimental lab. The participants are informed that: “*This urn contains new prizes. One such prize is the one you see. The urn contains no prizes similar to what you have been observing as a result of random draws from the other urn.*” The experimental assistant subsequently brings in the new urn and draws one ball from it (D_1^n in Figure 1), revealing one new prize s to the participants. We do not reveal any other information about the contents of the new urn. As before, the value of the new prize s varies with the prize condition of the session. In the *low* prize condition, the new urn contains 15 balls labeled ‘15’, while in the *high* prize condition it contains 15 balls labeled ‘375’. We then proceed to empty the contents of the new urn into the original urn, leading to the *updated* urn. The experimental assistant subsequently makes a draw from the updated urn (D_1^u in Figure 1). The outcome of the draw from the updated urn is used to determine the potential payment for the report of WTA^o . However, the draw is concealed from the participants immediately after it is made. As for *IS*, it is revealed only at the very end of the experiment when the final payment to the participants is displayed, provided that the WTA^o

report is selected for payment. Thus, in contrast to the *IS* condition where the draw determining payment for the report of WTA^o is made before the urn is updated, the draw in the *PS* condition is made from the updated urn: participants were forewarned that this may happen.

The participants are then again asked to estimate the probability of each prize ($p_{80}^u, p_{190}^u, p_s^u$) and to state their willingness-to-accept for the prospect to draw a prize from the urn (WTA^u). Following these reports, the experimental assistant draws a ball from the updated urn (D_2^u in Figure 1). This draw is used to determine the payment to the respondents, provided the WTA^u report is selected for payment at the very end of the experiment. Similarly to *IS*, this draw is concealed from the respondents and only revealed at the very end of the experiment if this decision is selected for payment. Thus, the last two stages of the *PS* condition coincide with the last two stages of the *IS* condition (stages 6 and 7 in Figure 1).

The payment for the urn tasks is determined as follows. One item is randomly selected from the set of all reported probability estimates and the two $WTAs$. This item is played out and the resultant payoff is added to a participant's payment. We incentivize the reported probability estimates according to the Karni (2009) method.⁶ The reports of WTA are incentivized using the BDM procedure (Becker et al., 1964). Both mechanisms induce truth telling and are robust to varying risk attitudes.

The BDM procedure works as follows: If one of the $WTAs$ is selected to be payment-relevant, the computer draws a random price for the prospect between the smallest and largest prizes in the urn.⁷ If the realization of the random price is such that the reported WTA exceeds the randomly generated price, the participant keeps the prospect and the payoff is determined by the hidden draw made during the experiment from the original urn in *IS* condition for WTA^o (see D_{21}^o in Figure 1); from the updated urn for WTA^u (see D_1^u in Figure 1); and from the updated urn for both WTA^o and WTA^u in *PS* condition (see D_1^u and D_2^u in Figure 1). If WTA is smaller than the random price, the participant sells the prospect, and receives the random price. The Karni mechanism works similarly for probabilities. A random probability (between 0 and 1) is drawn. If the reported probability estimate exceeds this randomly drawn probability, the participant is rewarded with the relevant prize

⁶See the first page of the experimental instructions in Appendix A following the heading "Likelihoods of events – Reporting and Earnings" for more details on how this was explained to the participants as well as for further details on the method itself.

⁷Similarly to Isoni et al. (2011), these bounds are not communicated to the participants.

according to the actual probability of that prize in the urn and gets nothing with the complementary probability. Alternatively, the participant is rewarded with the relevant prize according to the randomly drawn probability and gets nothing with the complementary probability.

The objective of our study is to examine whether subjects expect the unexpected and to investigate how their beliefs change after encountering a new event. For a new event to be unexpected, it should be unannounced and/or ruled out. To elicit updated beliefs within our design in an incentive compatible fashion, the new event must have direct and immediate payment consequences. Ideally, our treatment would have both of these characteristics. However, if our individual conditions had both of these characteristics, we could be accused of deception for explicitly or implicitly signaling that there would be no new event but then implementing the latter and making it payment-relevant. In light of this constraint, we designed our experiment to have two conditions, each of which has one and only one of these characteristics. Notice that the aim of the two conditions, *IS* and *PS*, is not to contrast belief updating directly between the two, but rather, to study belief updating in two closely related situations. In *IS*, the new event is unannounced while in *PS* the new event is payment-relevant because the payment is determined by the updated urn. Specifically, in contrast to *IS*, the respondents in *PS* are forewarned that new prizes may be added to the urn (thus, making the new event potentially foreseeable). Moreover, while in *IS* the first (potentially) payment-relevant draw (D_{21}^o in Figure 1) is made from the original urn, the corresponding draw in *PS* is made from the updated urn (D_1^u in Figure 1). Our approach also allows us to test the robustness of results regarding reverse Bayesianism across different settings.

In addition to these two differences, the conditions *IS* and *PS* differ along two other dimensions. These two differences were implemented to make the new event in *IS* as unexpected as possible and to avoid misleading the respondents in *PS* as much as possible. These differences pertain to the information about the compositions of the original and new urns, respectively, under the two conditions. Unlike in the *IS* condition, we did not tell the respondents in the *PS* condition that the original urn contains two and only two prizes. Otherwise, one could argue that we are sending a message that contradicts the possibility that the content of the urn may be changed or that we are trying to mislead the respondents. Even after making 20 draws from the original urn, the respondents in *PS* may expect to encounter a prize value that they

have not yet observed. Thus, the possibility of drawing a new prize is conceivable in *PS*. Finally, the fourth difference is consonant with the third. The respondents in *IS* are told that the new urn contains only balls with the newly revealed prize (D_1^n in Figure 1). In contrast, under *PS* the possibility that the new urn may contain prizes other than the newly revealed prize is not ruled out.

Once the urn task is completed, we elicit risk preferences using an incentivized Eckel and Grossman (2008) task. In this task, individuals pick one lottery from a set of binary lotteries. The lotteries in the choice set vary in terms of their expected values and variances with the chosen lottery revealing a participant’s risk attitude. The lotteries chosen by the participants are “played out” at the end of the experiment and the earnings for these choices are added to the rest of the earnings of each participant.

Following the elicitation of risk preferences, the participants complete a short Raven Advanced Progressive Matrices (APM) test (Raven et al., 1998b,a). Raven’s Progressive Matrices provide an effective non-verbal avenue to measure reasoning and general cognitive ability. In order to shorten the duration of this test, we follow Bors and Stokes (1998) in using 12 from the total of 36 matrices from Set II of the APM. Matrices from Set II of the APM are appropriate for adults and adolescents of higher than average intelligence. Participants are allowed a maximum of 10 minutes. The participants are informed that two of these 12 matrices are selected at random for payment and that they will receive €1 for each correct choice. The sessions are concluded with some general demographic questions and a final screen informing the participants about their total earnings.

We include the experimental instructions in Appendix A. Within this Appendix in table 14, we also include a table that lists the draw realizations across all sessions and the average WTA reported by session. The design was pre-registered at the AEA RCT Registry <https://www.socialscienceregistry.org/trials/3815>.

Implementation

Experiment 1 was conducted at the Alfred-Weber-Institute Experimental Lab at the University of Heidelberg and the Karlsruhe Decision & Design Lab (*KD²Lab*) at the Karlsruhe Institute of Technology. Conducting the experiment in two separate locations was done to reduce the possibility that former participants communicate details about the experiment to later participants, and given the nature of this experiment

it would have been particularly problematic. The recruitment of participants took place via SONA systems for Heidelberg and ORSEE (Greiner, 2015) for Karlsruhe. A total of 344 participants participated in the experimental sessions.⁸ The participants earned an average of €18.4, including a show-up fee of €4. The software used for the entire experiment was z-Tree (Fischbacher, 2007). The ethical approval for this design was granted by the Humanities and Social Sciences Research Ethics Sub-Co at the University of Warwick under DRAW Umbrella Approval (Ref: HSS 49/18-19, DR@W submission ID: 485613261).

3.2 Results

3.2.1 Reverse Bayesianism

We start by testing whether belief updating is consistent with reverse Bayesianism. In our framework, reverse Bayesianism requires that the elicited probability ratios of the prizes in the original urn, namely the prizes of 80 and 190 tokens, remain unchanged after the original urn is updated. Formally, Hypothesis 1 for this experiment implies:

$$\Delta R = \frac{\hat{P}_{80}^u}{\hat{P}_{190}^u} - \frac{\hat{P}_{80}^o}{\hat{P}_{190}^o} = 0. \quad (1)$$

The information provided to the participants in all four treatments unambiguously reveals that the number of balls worth 80 tokens and the number of balls worth 190 tokens remain unchanged after the urn is updated. For *(IS, low prize)* and *(IS, high prize)* treatments, we informed the participants that the new urn contains only the newly revealed prize. For *(PS, low prize)* and *(PS, high prize)* treatments, we told them that the new urn contains no prizes similar to what they have been observing as a result of random draws from the original urn. Thus, if a participant’s beliefs are given by a singleton probability distribution, i.e., she is probabilistically sophisticated (Machina and Schmeidler, 1992), throughout the experiment, then $\Delta R = 0$ as long as that participant forms her beliefs on all of the information provided in the experimental instructions.

Table 1 contains the results of Wilcoxon signed-rank tests for all four treatments.

⁸We had 234 participants in the sessions at Heidelberg and 110 participants in the sessions at Karlsruhe. In Heidelberg, 46 participants were in the *(IS, low prize)* treatment, 58 in *(IS, high prize)*, 59 in *(PS, low prize)*, and 71 in *(IS, high prize)*. In Karlsruhe, the numbers were 30, 17, 34, and 29, respectively. Data is not qualitatively different across the two subject pools.

We fail to reject the hypothesis in three out of the four treatments, both before and after correcting for multiple testing. For these three treatments, we indeed find a precisely estimated null effect.⁹ Only in one treatment, (*PS, high prize*), we reject the null hypothesis of reverse Bayesianism. Looking at the confidence intervals derived from a t-test, we note that 95% of participants deviate very little from 0, even in the treatment where we reject the null hypothesis. Subjects seem to change the probability ratio in the updated urn by decreasing \hat{p}_{190} slightly more than \hat{p}_{80} . Since the original urn contains a larger number of balls worth 190 tokens than balls worth 80 tokens, this adjustment of reported probabilities might arise from subjects feeling more comfortable with decreasing a larger number. We report a similar behavioral pattern in experiment 2.

It is important to highlight that our data analysis so far fails to reject the null hypothesis not because of limited power, but because the null is supported by our data. To provide further evidence in support for the predicted null effect of reverse Bayesianism, we rely on Bayesian inference statistics.¹⁰ Specifically, we implement the JZS test developed by Rouder et al. (2009), which is a Bayesian alternative to t-tests. Like other Bayesian methods it offers a researcher the possibility to state whether the data contains evidence in support of the null hypothesis. Crucially, the JZS test makes assumptions about the prior distributions of the effect size and variance, thus circumventing the problem of Bayes factors favoring the null hypothesis when a non-informative prior is used for the alternative.¹¹ The usual rule-of-thumb for interpreting Bayes factors applied to the JZS test is that a factor of 0 provides strong evidence for the alternative, 1 is inconclusive (the predictions of the null and the alternative cannot be disentangled) and values above 3 provide strong evidence for the null hypothesis (here reverse Bayesianism). The last column in Table 1 reports the Bayes factors. These values are consistently above 3 in all four treatments, thus

⁹Testing the ratios before and after the update also shows that participants do not simply provide equal estimates for both prizes, that is having ratios equal to 1. On average, ratios before and after the update are smaller than 1, with $p < 0.001$, Wilcoxon signed-rank test.

¹⁰For discussions and examples of the use of Bayesian inference statistics in the social sciences, see, e.g., Bayarri et al. (2016); Dienes (2011).

¹¹Specifically, the JZS test assumes that the null hypothesis is a point with $H_0 : \delta = 0$, while the alternative effect size $\delta = \frac{\mu}{\sigma}$ follows a Cauchy distribution (which assumes that more extreme values are more unlikely) with $H_1 : \delta \sim Cauchy(r) = 1$ (where r is a scaling parameter). The prior for the variance is given by $p(\sigma^2) \propto \frac{1}{\sigma^2}$. This prior is deliberately non-informative, as the variance is relevant for both hypotheses.

offering strong evidence in support of reverse Bayesianism.¹²

Table 1: Average ratio changes before and after the urn is updated.

		Obs	Avg ratio change	p-value	p-value (corr)	95%CI	Bayes factor
IS	low prize	75	0.007	0.375	1.000	[−0.06, 0.05]	14.76
	high prize	75	−0.039	0.981	1.000	[−0.06, 0.14]	6.57
PS	low prize	93	0.016	0.918	1.000	[−0.06, 0.03]	9.72
	high prize	100	−0.007	0.011	0.043	[−0.04, 0.05]	16.35

Wilcoxon signed-rank test, p-values corrected by Bonferroni-Holm procedure, confidence interval from one sample t-test, Bayes factor from JZS test.

The ratio of the probabilities of the prizes observed by the participants may remain unchanged under two scenarios. First, participants may simply not change their estimate for any of the previously observed prizes. Table 2, column 7, shows that this is rarely the case in any of the treatments. Second, participants may change their estimates in a way that keeps the ratios constant. Notice that this is far from trivial for many constellations of estimates. Although we observe such instances (column 4), overall this is considerably less prevalent than the instances when participants either increased or decreased the ratio. Figure 2 shows the distribution of ratio changes both for all participants and for all instances where the participants adjusted all their estimates after the update. The data is pooled over all four treatments.¹³ Our analysis reveals that the absolute differences of the ratios tend to be concentrated around zero. That is, the change in the ratio is relatively small even for those participants who updated their beliefs to a different ratio. However even when the ratio changes, as we already noted from Table 1, most participants who change their ratios do so only slightly, and in no systematic direction. Thus, the overall null effect in support of reverse Bayesianism materializes because the changes in the ratio of the probabilities tend to be very small.

The evidence we have provided so far indicates that on average we observe no changes to the ratios. It is important to emphasize that this does not mean that the participants are passive and do not update their estimates after the original urn is updated. In Table 3, we test whether the estimates for known outcomes are

¹²Juxtaposing the estimated confidence interval and the Bayes factors for the (*PS, high prize*) treatment suggests that the rejection of the null according to the Wilcoxon signed-rank test is possibly driven by a relatively large proportion of positive changes that are nevertheless of very small magnitude. This is further supported by the bottom right panel of Figure A3 in Appendix D.

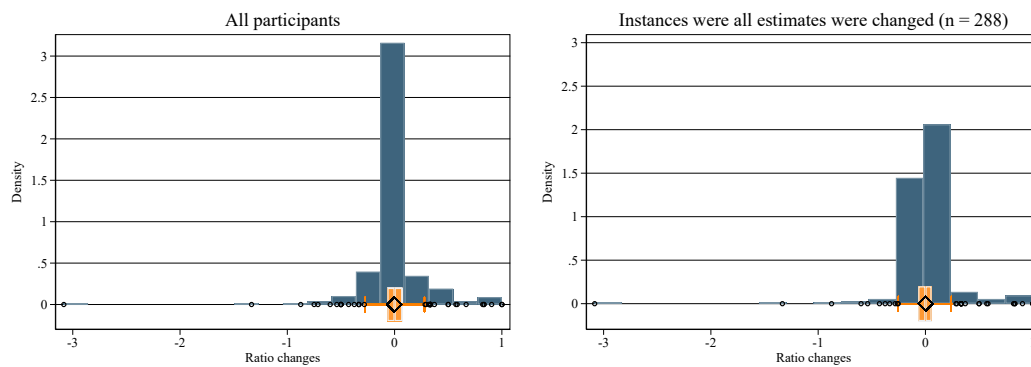
¹³See Figure A3 in Appendix D for the equivalent distributions for each treatment separately.

Table 2: Changes of the probability ratios following the update of the urn.

		Increased	Decreased	Const ratio	p-value	p-value (corr)	Unchanged Est
IS	low prize	29	23	23	0.488	1.000	1
	high prize	31	32	12	1.000	1.000	1
PS	low prize	33	37	23	0.720	1.000	0
	high prize	29	61	10	0.001	0.004	4

Matched pairs sign test, p-values corrected by Bonferroni-Holm procedure. 'Unchanged Est.' denotes the subset of those holding their ratios constant while not changing any of their estimates.

Figure 2: Histograms of the changes in the ratios following the update of the urn.



Histogram in blue, box plot in orange, outliers (circles) and mean (diamond) in black.

significantly different between the original and updated urns. Specifically, we compare individual estimates of the prizes before and after the update. We consistently find that the estimates of the known outcomes are significantly higher in the original urn as compared to the updated urn. This indicates that the participants do react to the updating of the urn when reporting their proportion estimates. Summarizing:

Exp.1 - Result 1. *The participants hold their ratios approximately constant after encountering an unexpected event. Thus, they update their beliefs according to reverse Bayesianism (Hypothesis 1).*

Moreover, we do not find a statistically significant relation between cognitive ability and ratio differences (see Table A1 in Appendix D). That is, behavior is very similar for high and low cognitive ability participants, thus, unlikely to be due to errors caused by lack of understanding.

Exp.1 - Result 2. *Cognitive ability has no mediating effect on the deviations from reverse Bayesianism.*

Table 3: Changes of known outcome estimates after observing the update of the urn.

		Obs	Diff	p-value	p-value (corr)
IS, low prize	$\hat{p}_{80}^u - \hat{p}_{80}^o$	76	-0.101	0.000	0.000
	$\hat{p}_{190}^u - \hat{p}_{190}^o$	76	-0.130	0.000	0.000
IS, high prize	$\hat{p}_{80}^u - \hat{p}_{80}^o$	75	-0.102	0.000	0.000
	$\hat{p}_{190}^u - \hat{p}_{190}^o$	75	-0.125	0.000	0.000
PS, low prize	$\hat{p}_{80}^u - \hat{p}_{80}^o$	93	-0.100	0.000	0.000
	$\hat{p}_{190}^u - \hat{p}_{190}^o$	93	-0.136	0.000	0.000
PS, high prize	$\hat{p}_{80}^u - \hat{p}_{80}^o$	100	-0.075	0.000	0.000
	$\hat{p}_{190}^u - \hat{p}_{190}^o$	100	-0.108	0.000	0.000

Wilcoxon signed-rank test, p-values corrected by Bonferroni-Holm procedure.

3.2.2 Residuals and Valuations

The experiment yields a number of additional interesting findings. First, we focus on the estimates of residual probabilities provided by the participants and test Hypotheses 2 and 3. Turning to Hypothesis 2, the participants in the *IS* condition were

informed that the original urn contains only two possible prizes, and that in the updated urn only one extra prize was added. If the respondents took this information into account, the set of residual consequences x , as defined in Section 2, would be empty, thus $\hat{p}_x^o \neq 0$ and $\hat{p}_x^u \neq 0$ would likely only materialize due to individual idiosyncratic errors. In contrast, since in the *PS* condition we informed the participants about the possibility of adding new prizes and we did not state that the updated urn contains only one new prize, $\hat{p}_x^o > 0$ and $\hat{p}_x^u > 0$ could occur as a result of an expectation of unknown events. Thus, one could expect that the respondents in the *PS* condition are more likely to assign a strictly positive probability to encountering a prize that they have not seen before.

Table 4 shows that the hypothesis $\hat{p}_x^t = 0$ (t is either o for original or u for updated) cannot be rejected for the *IS* condition, both for the original and for the updated urn, thus giving support to Hypothesis 2 (see also Figures A1 and A2 in Appendix D for a graphical representation of the residual estimates). Even in the *PS* condition, the hypothesis $\hat{p}_x^t = 0$ cannot be rejected for the original urn. Moreover, the hypothesis cannot be rejected in the *PS* condition for the updated urn in the case of (a perhaps less salient) high prize surprise. The hypothesis that $\hat{p}_x^t = 0$ is rejected in the updated urn in the (*PS*, *low prize*) treatment. Moreover, column 2 of Table 4 reveals that in the *PS* condition, $\hat{p}_x^u > 0$ for a larger number of participants compared to other conditions. It seems plausible that at the point in the experiment where the participants have only seen the original urn, the event that the urn may be updated is unforeseeable to many of them. In contrast, the latter event is unlikely to be unforeseeable after the participants have witnessed the update of the original urn.

Exp.1 - Result 3. *Overall, the participants do not expect unforeseen events, thus, lending support to the first part of Hypothesis 2. Evidence on the second part of Hypothesis 2 is mixed: after encountering adverse new events, participants anticipate unforeseen events, thus, supporting the second part. However, after favorable unforeseen events, they do not, thus rejecting the second part.*

We now turn to Hypothesis 3. Following the same reasoning as before, we expect some awareness for *PS* but not for *IS*, and test for changes in the residuals after receiving new information. Table 5 shows that the hypothesis that $\hat{p}_x^o = \hat{p}_x^u$ cannot be rejected in the *IS* treatments. However, there is again some evidence for $\hat{p}_x^o(PS) \neq \hat{p}_x^u(PS)$, when the surprising event entails a low prize.

Table 4: Residuals different from 0.

		$\hat{p}_x^t = 0$	$\hat{p}_x^t > 0$	$\hat{p}_x^t < 0$	p-value	p-value (corr)
IS, original	low prize	74	1	1	0.993	1.000
	high prize	71	3	1	0.314	1.000
PS, original	low prize	92	0	1	0.317	1.000
	high prize	90	6	4	0.549	1.000
IS, updated	low prize	61	10	5	0.251	1.000
	high prize	65	7	3	0.228	1.000
PS, updated	low prize	74	16	3	0.004	0.028
	high prize	84	11	5	0.146	1.000

Wilcoxon signed-rank test, p-values corrected by Bonferroni-Holm procedure.

Exp.1 - Result 4. *With the exception of adverse new event for PS condition, the participants do not adjust their residual beliefs following an unforeseen event, thus, not rejecting Hypothesis 3.*

Table 5: Differences between the residual before and after the surprise: $\Delta\hat{p}_x = \hat{p}_x^u - \hat{p}_x^o$

		$\Delta\hat{p}_x = 0$	$\Delta\hat{p}_x > 0$	$\Delta\hat{p}_x < 0$	p-value	p-value (corr)
IS	low prize	60	11	5	0.173	0.692
	high prize	63	6	6	0.937	1.000
PS	low prize	73	17	3	0.002	0.009
	high prize	82	11	7	0.345	1.000

Wilcoxon signed-rank test, p-values corrected by Bonferroni-Holm procedure.

Given our design, awareness of encountering an unforeseen event and the way an unforeseen event is experienced will affect the stated *WTAs*. It follows from Karni and Vierø (2013, 2017) that two key factors are at play in the participants' evaluation of uncertain prospects. The first characteristic pertains to the participants' updated beliefs; how much of the probability weight is shifted from the known prizes to the newly observed and yet unobserved prizes. The second concerns the participants' attitude towards the unknown; whether and how much they like or dislike the unknown. To determine the relative importance of these channels, we compare the elicited willingness to accept measures before and after the urns are updated, for both the *IS* and *PS* conditions and both levels of the new prize. Table 6 reveals that $WTA^o(PS) > WTA^o(IS)$ in both high and low prize conditions. As one could expect, the *WTAs* for the updated urn are lower for the low prize than for the high prize,

and again the *PS* condition elicits higher valuations; the latter effect is, however, not significant.

The regression analysis with controls for gender, cognitive ability, degree of risk aversion and the observed sample, confirms the effect of the *PS* treatment and of the high prize in the updated urn *WTA* (see Tables A2 and A3 in Appendix D). Overall, it seems that the more uncertain situation in condition *PS* elicits higher valuations. That is, in the context of unforeseen events, *hope* seems to dominate *fear* (Viscusi and Chesson, 1999). Additionally, it seems that the belief about the number of 190 prizes in the urn is an important driver of the *WTA*, not the actually observed number. As a caveat, we note that *WTA* measurement in the context of uncertainty and ambiguity has been found to elicit relatively higher valuations for more uncertain prospects (Trautmann et al., 2011; Trautmann and Schmidt, 2012). The selling-price context seems to induce decision makers to focus on the potentially forgone benefits from selling a highly uncertain prospect. This effect re-emerges here.

Exp.1 - Result 5. *The increased uncertainty in condition PS results in higher valuations of the urn: the participants appear to view the unknown with hope rather than with fear.*

Table 6: *WTA* for a draw from the urn by treatment.

Original urn: <i>WTA</i>^o					
	IS	PS	Diff	p-value	p-value (corr)
Low prize	110.39	138.47	-28.08	0.008	0.031
High prize	110.48	134.81	-24.33	0.002	0.006

Wilcoxon signed-rank test, p-values corrected by Bonferroni-Holm procedure.

Updated urn: <i>WTA</i>^u					
	IS	PS	Diff	p-value	p-value (corr)
Low prize	86.45	96.70	-10.25	0.074	0.295
High prize	153.53	178.25	-24.72	0.160	0.639

Wilcoxon signed-rank test, p-values corrected by Bonferroni-Holm procedure.

4 Experiment 2

4.1 Design

Experiment 2 tests how individuals perceive uncertainty and update their beliefs in light of new events which are foreseeable, but potentially unforeseen. Each participant in Experiment 2 individually draws a sample of 30 colored marbles with replacement from a virtual urn in each of four different tasks. The urns in each of the four tasks contain a total of 100 virtual marbles, which is known to the participants. No other information about the composition of the urns is revealed to the participants prior to them making the draws from the urns. There are two types of potentially unforeseen events in this design. The first type entails encountering a new color, while the second pertains to observing some *specific* new color. The way the task is set up, we expect participants to assign a non-zero belief to the first type of event. This, however, does not mean that participants cannot be surprised in this experiment. The belief participants assign to specific events of the second type could vary considerably between participants, depending on their imagination (Shackle, 1949). Hence, while we expect participants to be aware about the existence of new colors in general, they should be surprised by *specific* colors unforeseen by them. In our investigation, we focus on eliciting beliefs about the first type of events, which is also the union of all possible events of the second type.¹⁴

Each participant is randomly allocated to one of two treatments. In the *two colors* treatment, the urn in the first task contains only two colors. In the *four colors* treatment, the urn in the first task contains four colors in total. The purpose of this design is to test if encountering a larger number of different colors in the first task increases their awareness that further surprises, in the form of new colors, might be possible in subsequent tasks. The compositions of the urns in the second, third and fourth tasks are the same across the two treatments. The urn in the second task contains three colors, the urn in the third task contains two colors and the urn in the fourth task contains four colors. Table 7 provides information on the exact compositions of the urns in the four tasks.

The urns with three and four colors contained a comparatively small number of some of the colors. This ensured that the likelihood of encountering a new and

¹⁴Exact colors for each task are randomized at the participant level. As a result, each participant observed a different sequence of colored marbles for each task.

Table 7: Numbers of different colors in the tasks.

	Task 1		Task 2	Task 3	Task 4
	Two colors	Four colors			
Color 1	55	40	53	75	48
Color 2	45	28	35	25	28
Color 3		20	12		12
Color 4		12			12

surprising outcome even after sampling several times was relatively large for these urns.

After each sample draw, the drawn marble is presented on a participant’s screen, both with an image of a marble of that specific color and the name of the color. In addition, the outcomes of all previous draws are depicted in a small caption at the bottom of the participants’ screens. This provides the participants with a full overview of the past draws and mitigates the effects of memory limitations.

After each draw, the participants are asked to report their estimates of the contents of the urn. Specifically, they are asked to state their estimate of (i) the number of marbles of the color they *just drew*; (ii) the number of marbles of *each other* color they have previously drawn in this task; and (iii) the number of marbles of *yet unobserved* colors.

The third item provides us with a residual probability assigned by the respondents to any conceivable color not yet observed during the draws. Thus, in contrast to Experiment 1, the residual probability is elicited explicitly in Experiment 2. The participants submit their estimates by entering integer numbers between 0 and 100 into form fields on their screens. We provide them with one individual form field for each color drawn up to that point, as well as with one form field for their estimate of the number of marbles of yet unobserved colors. This design allows us to trace how the respondents’ estimates and ratios of these estimates are adjusted once unforeseen information becomes available. In addition, we can track how the estimates of likelihoods of yet unobserved outcomes evolve over the sampling process. To make the submission of estimates easier, we additionally provide respondents with buttons that allow them to fill in their last estimates for a color and “plus” and “minus” buttons to increase/decrease these estimates by one per click.¹⁵

¹⁵An example screenshot of the sampling screen is included in Appendix C.

We again use the Karni (2009) method to ensure that the participants are incentivized to submit their estimates truthfully. At the end of the experiment, for each of the four tasks, one of the 30 sample draws is randomly selected and one item from the set of reported estimates for that draw is also randomly chosen by the computer (this could involve an estimate of yet unobserved colors). The payment mechanism is implemented for that reported estimate. Participants can earn £6 or nothing from each task depending on the outcome according to the incentivization method.¹⁶

Following the four sample tasks, the participants complete the same incentivized APM task described used in Experiment 1 (see Section 3.1). After the APM, participants are informed in detail about their total earnings. The sessions are concluded with a questionnaire containing demographic questions and a question eliciting risk attitudes.

In addition to the hypotheses presented in Section 2, we additionally test the following:

Hypothesis 4.

- (a) *The probability \hat{p}_x^t assigned each round to yet unobserved outcomes decreases with the draws (t) made from an urn:*

$$t \rightarrow 30 \Rightarrow \hat{p}_x^t \downarrow$$

- (b) *The probability assigned to yet unobserved outcomes in tasks 2, 3 and 4 decreases faster for the participants in the two colors treatment than for the participants in the four colors treatment.*

Hypothesis 4 tests if the participants learn to expect fewer surprises towards the end of the sampling process, as well as whether this process might vary in terms of speed across treatment conditions. Specifically, participants in the *four colors* treatment might be more aware that more colors are possible.

Following some recent insights on the link between cognitive ability and rational behavior (e.g. Alaoui and Penta, 2016; Gill and Prowse, 2016) we also test two hypotheses related to cognitive ability.

¹⁶See the third page of the instructions for this experiment in Appendix B under the heading “Getting paid for good predictions” for information on how this was explained to the participants as well as for further details on the method itself.

Hypothesis 5. *The participants with higher cognitive ability exhibit fewer deviations from reverse-Bayesianism.*

Hypothesis 6. *The participants with higher cognitive ability expect yet unobserved outcomes up to a later point in the sampling process than the participants with lower cognitive ability.*

We include the experimental instructions in Appendix B. The design was pre-registered at the AEA RCT Registry <https://www.socialscienceregistry.org/trials/5499>.

Implementation

Experiment 2 took place at the Behavioral Science Lab at the University of Warwick. The recruitment was conducted with the DRAW (Decision Research at Warwick) system, based on the SONA systems. A total of 174 individuals participated in the experiment, 89 in the *two colors* treatment and 85 in the *four colors* treatment. Note that we originally intended to have 150 participants in each treatment (as specified in our pre-registration). However, due to the unforeseen onset of the COVID-19 pandemic we were not able to gather additional data. The average payment was £16.87, including a show-up fee of £3. The software for the experiment was programmed in otree (Chen et al., 2016). Ethical Approval for this design was granted by the Humanities and Social Sciences Research Ethics Sub-Co at the University of Warwick under DRAW Umbrella Approval (Ref: HSSREC 104/19-20, DR@W submission ID: 514470520).

4.2 Results

4.2.1 Reverse Bayesianism

We first test whether behavior is consistent with reverse Bayesianism, which corresponds to Hypothesis 1. We examine the difference in the ratios of previously observed colors directly before and directly after observing a new color. The ratios are defined for pairs of colors that have already been observed and on the basis of the relative magnitudes of the estimated likelihoods of these two colors immediately before a new color is observed. Specifically, we define \hat{p}_H^o as the estimate of the likelihood for the color that is considered by a participant to be more likely and \hat{p}_L^o as the estimate

for the color that is considered to be less likely. Both of these estimates are for the sample draw right before the third color is observed for the first time. We also define \hat{p}_H^u and \hat{p}_L^u as the estimates of the likelihoods of these two colors immediately after the third color is first observed. Specifically, we test:

$$\Delta R^3 = \frac{\hat{p}_H^u}{\hat{p}_L^u} - \frac{\hat{p}_H^o}{\hat{p}_L^o} = 0.$$

For the belief update following the observation of a fourth color (having already seen three colors), we have three ratios to consider. Define \hat{p}_H^o , \hat{p}_M^o and \hat{p}_L^o as the estimates for the color considered most likely, second most likely, and least likely, of the three colors that have already been observed, in the sample draw right before the fourth color is observed for the first time. We also let \hat{p}_H^u , \hat{p}_M^u and \hat{p}_L^u denote the respective estimates for these three colors in the sample draw immediately after the fourth color is first observed. We test three relationships:

$$\Delta R_1^4 = \frac{\hat{p}_H^u}{\hat{p}_M^u} - \frac{\hat{p}_H^o}{\hat{p}_M^o} = 0, \quad \Delta R_2^4 = \frac{\hat{p}_M^u}{\hat{p}_L^u} - \frac{\hat{p}_M^o}{\hat{p}_L^o} = 0, \quad \Delta R_3^4 = \frac{\hat{p}_H^u}{\hat{p}_L^u} - \frac{\hat{p}_H^o}{\hat{p}_L^o} = 0.$$

Table 8 contains the results of Wilcoxon signed-rank tests for all ratio changes (as described just above) including the data from both treatments.¹⁷ For Tasks 1 and 4, we also pool the three ratio changes after observing the fourth outcome (indicated in Table 8 as ΔR_P^4). The results indicate that for all eleven tests that we conduct, there is no significant change in the ratios after controlling for multiple tests.¹⁸ Even when not controlling for multiple testing, there is no statistically significant change in eight of the eleven tested ratios.¹⁹ Finally, from the 95% confidence intervals (obtained from t-tests) in column 5, we notice that the ratios do not change substantially. This is especially manifest when the urn contains three colors so that the participants have to keep only one ratio constant, denoted as ΔR^3 in the table. Similarly to the analysis for Experiment 1, we also include the Bayes factors in the last column of Table 8. Except for three cases, the Bayes factors are above 3. In two of the three cases, however, the confidence intervals still include 0. Table 8 thus provides strong evidence in support of reverse Bayesianism.

¹⁷Our results did not differ across treatments, thus our tables pool treatments from here on.

¹⁸The urn in task 3 contains only two colors. Hence, the third outcome surprise is not possible and, consequently, we do not analyze it in this case.

¹⁹As in Experiment 1, we also test whether participants before and after the update simply provide equal estimates for both prizes, that is having ratios equal to 1. On average, ratios before and after the update are larger than 1, with $p < 0.001$, Wilcoxon signed-rank test.

It is worth noticing that all average ratio changes, albeit statistically insignificant, are negative. This suggests the possibility of a minor pattern that is not captured because of the variability in the data. As in experiment 1, subjects might be decreasing the likelihoods of events to which they originally assigned relatively high probabilities by relatively large amounts (in this case, \hat{p}_H decreases slightly more than \hat{p}_M that, in turn, decreases slightly more than \hat{p}_L). This effect, if it exists, is small in magnitude and it remains insignificant after correcting for multiple testing and even when we pool together all of the changes, including the changes following the fourth event (i.e. ΔR_P^4). Furthermore, the fact that the absolute value ΔR_P^4 in task 1 is larger than in task 4 hints at a potential learning effect, where the participants' behavior aligns more with reverse Bayesianism in later tasks.

Table 8: Average ratio changes before vs. after observing a new color.

		Obs	Avg ratio change	p-value	p-value (corr)	95%CI	Bayes factor
Task 1	ΔR^3	85	-1.365	0.172	1.000	[-0.10, 0.29]	5.32
	ΔR_1^4	84	-0.548	0.584	1.000	[-0.79, 0.22]	4.44
	ΔR_2^4	84	-2.134	0.033	0.362	[-1.27, 0.04]	1.58
	ΔR_3^4	84	-1.005	0.315	1.000	[-0.52, 0.33]	7.52
Pooled	ΔR_P^4	252	-2.229	0.026	0.284	[-0.64, -0.03]	1.49
Task 2	ΔR^3	169	-2.632	0.008	0.093	[-0.31, 0.01]	2.26
Task 4	ΔR^3	173	-0.648	0.517	1.000	[-0.26, 0.06]	5.71
	ΔR_1^4	164	-0.048	0.962	1.000	[-0.07, 0.25]	6.05
	ΔR_2^4	163	-0.067	0.946	1.000	[-0.19, 0.69]	6.09
	ΔR_3^4	163	-0.148	0.883	1.000	[-0.14, 0.27]	9.46
Pooled	ΔR_P^4	490	-0.203	0.839	1.000	[-0.03, 0.30]	5.64

Wilcoxon signed-rank test, p-values corrected by Bonferroni-Holm procedure, confidence interval from one sample t-test, Bayes factor from JZS test.

As noted in the analysis for Experiment 1, the finding that the ratios do not significantly change in aggregate may derive from (i) participants holding their estimates unchanged, (ii) participants holding the ratios constant, or (iii) some participants increasing while some decreasing their ratios, with the effects canceling each other out on average. Table 9 indicates that, in contrast to Experiment 1, a large number of participants hold their estimates unchanged. A possible explanation for the participants in Experiment 2 not changing their estimates is the provision of a button to fill-in their previous estimate to simplify the dynamic task for the participants. However, there is also a substantial number of participants who hold their ratio constant, while adjusting the separate probability estimates. Again, holding the ratio constant while

adjusting the separate estimates, especially after the fourth color is observed, is far from trivial. It indicates that participants actively aim to keep their ratios constant even when changing their separate estimates. At the same time, a substantial share of the participants change their ratios, but there is no systematic effect in the way they do it: after correcting for multiple testing, there are no significant differences in the deviations from constant ratios in the increasing or decreasing directions; seven of the nine uncorrected tests support constant average ratios.²⁰

Table 9: Changes of the ratios before vs. after observing a new outcome.

		Increased	Decreased	Const ratio	p-value	p-value (corr)	Unchanged Est
Task 1	ΔR^3	16	29	40	0.072	0.797	26
	ΔR_1^4	19	21	44	0.875	1.000	31
	ΔR_2^4	16	31	37	0.040	0.440	32
	ΔR_3^4	16	23	45	0.337	1.000	35
Pooled	ΔR_P^4	51	75	126	0.040	0.440	93
Task 2	ΔR^3	35	59	75	0.017	0.189	46
Task 4	ΔR^3	45	50	78	0.682	1.000	44
	ΔR_1^4	50	57	57	0.562	1.000	36
	ΔR_2^4	54	60	49	0.640	1.000	33
	ΔR_3^4	43	47	73	0.752	1.000	37
Pooled	ΔR_P^4	147	164	179	0.364	1.000	108

Matched pairs sign test, p-values corrected by Bonferroni-Holm procedure. 'Unchanged Est' denotes the subset of those holding their ratios constant while not changing any of their estimates.

This is further corroborated by Table 10, where we test whether the participants update their estimates of known outcomes after observing a new color. We again compare individual estimates of known outcomes before and after observing a new color. Similarly to Experiment 1, even though the ratios are on average constant, the individual estimates of known outcomes are updated downwards as new colors are observed.

Figure 3 illustrates that, as in Experiment 1, the ratio changes were closely concentrated around zero. This again holds both when we pool all participants in the experiment and when we pool all instances where all estimates are changed after observing a new color.

Exp.2 - Result 1. *The participants update their beliefs according to reverse Bayesianism, thus, providing support for Hypothesis 1.*

²⁰Again, we also find the above described (albeit still insignificant) pattern for the pooled ratios ΔR_P^4 in Tasks 1 and 4.

Once more, cognitive ability does not have a significant effect on the ratio deviations (Table A4 in Appendix E). Thus, there is no empirical evidence supporting Hypothesis 5. Importantly, the absolute changes in the ratios are also unaffected by the participants' expectations of further surprises, that is, whether they hold non-zero residual beliefs or not (Table A5 in Appendix E).

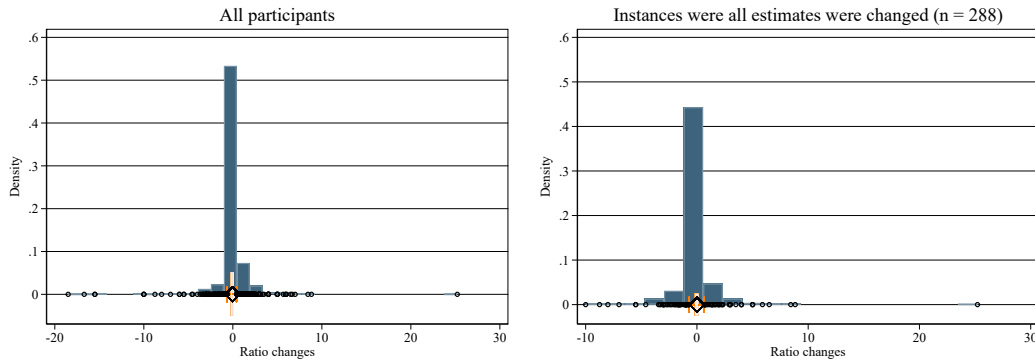
Exp.2 - Result 2. *Cognitive ability has no mediating effect on deviations from reverse Bayesianism, thus, we reject Hypothesis 5.*

Table 10: Changes of known outcome estimates after observing a new color.

		Obs	Diff	p-value	p-value (corr)
Task 1, after third color	$\hat{p}_H^u - \hat{p}_H^o$	85	-0.06	0.000	0.000
	$\hat{p}_L^u - \hat{p}_L^o$	85	-0.04	0.000	0.000
Task 1, after fourth color	$\hat{p}_H^u - \hat{p}_H^o$	84	-0.04	0.000	0.000
	$\hat{p}_M^u - \hat{p}_M^o$	84	-0.02	0.000	0.005
Task 2, after third color	$\hat{p}_L^u - \hat{p}_L^o$	84	-0.02	0.000	0.000
	$\hat{p}_H^u - \hat{p}_H^o$	169	-0.07	0.000	0.000
Task 4, after third color	$\hat{p}_H^u - \hat{p}_H^o$	169	-0.05	0.000	0.000
	$\hat{p}_L^u - \hat{p}_L^o$	174	-0.07	0.000	0.000
Task 4, after fourth color	$\hat{p}_H^u - \hat{p}_H^o$	164	-0.05	0.000	0.000
	$\hat{p}_M^u - \hat{p}_M^o$	164	-0.03	0.000	0.000
	$\hat{p}_L^u - \hat{p}_L^o$	164	-0.03	0.000	0.000

Wilcoxon signed-rank test, p-values corrected by Bonferroni-Holm procedure.

Figure 3: Histograms of the changes in the ratios following the update of the urn.



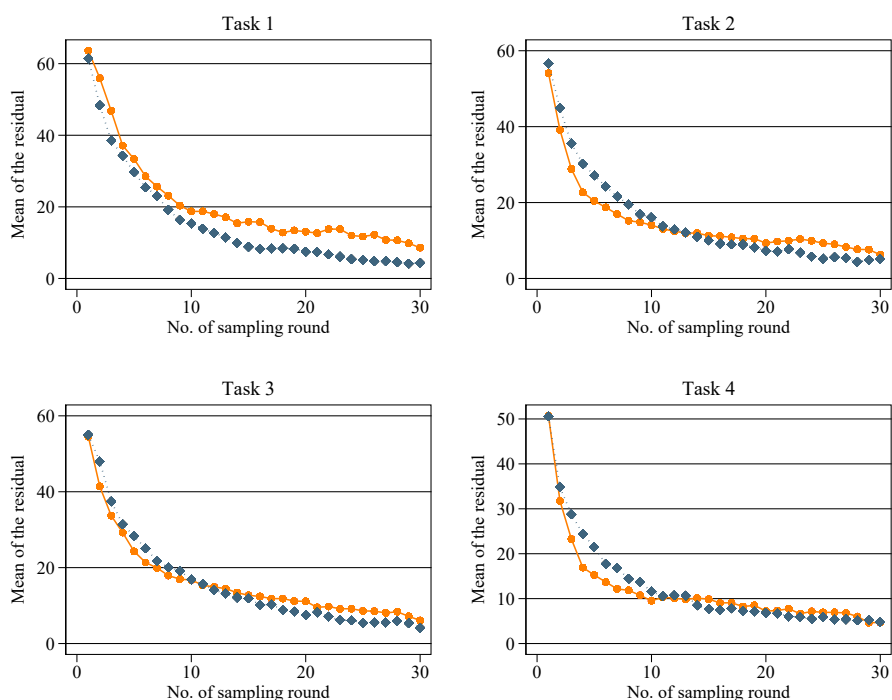
Histogram in blue, box plot in orange, outliers (circles) and mean (diamond) in black.

4.2.2 Analysis of Residuals and Belief Dynamics

We now turn to the estimates of the residual probabilities. Figure 4 depicts the evolution of the average residual \hat{p}_x^t over the 30 sample draws for each task and treatment. We find that for both treatments the average residual starts at a relatively high level and decreases quickly as more draws are made. However, even after the 30th sample draw, the average residual is well above 0. In Figure A4 in Appendix E, we present the distribution of these residuals; about one third of the participants expect further colors even after the 30th sample draw.

Exp.2 - Result 3. *On average, the participants anticipate unforeseen events, $\hat{p}_x^t > 0$, thus, providing support for Hypothesis 2.*

Figure 4: Dynamics of the residuals over the sampling process.



Note: The orange line depicts the residuals of *two outcomes* treatment while the blue line depicts the residuals of *four outcomes* treatment.

For the sake of exposition we postpone the analysis of Hypothesis 3, and consider Hypothesis 4 first. We already noted that the residuals on average monotonically decrease with sample draws (Figure 4). Simple correlations (across the whole set of participants) between the number of marbles already drawn from the virtual urn and

the stated residuals are negative for all tasks and across both treatments ($\rho < -0.311$, Pearson correlation coefficient), and thus support the first part of Hypothesis 4. However, there are no significant differences between the two treatments in the overall shape of the curve (Kolmogorov-Smirnov test, all p -values ≥ 0.994), thus rejecting the second part of Hypothesis 4 (treatment differences in awareness of encountering new colors). This indicates that participants do not adapt their estimation of residuals in later tasks in response to encountering more possible outcomes in the first task. This finding is intriguing. On the one hand, encountering more possible outcomes in the first task could raise the participants' awareness that the urns might contain more outcomes than initially expected, which was our prediction. On the other hand, as the participants have no information *ex ante* and as task 1 does not provide direct information on subsequent tasks, the null effect that we find might be perfectly rational.

Exp.2 - Result 4. *The residual probabilities decrease with the draws made from the urn, thus, supporting Hypothesis 4a. There is no difference in this trend between the two and four colors treatments, thus, we reject Hypothesis 4b.*

Turning to Hypothesis 3 (that the residual probability changes after observing new colors), there is a negative correlation between the number of colors already observed by the participants and their residual probabilities ($r_s < -0.272$, Spearman correlation coefficient). Examining the changes in the residuals directly before and after a new color is observed reveals the same picture. Table 11 depicts the residuals after each update, with \hat{p}_x^2 denoting the residual after the second color is observed, \hat{p}_x^3 after the third, and \hat{p}_x^4 after the fourth. There tends to be a significant drop in the subjective residual probability in 8 out of 9 instances. Thus, the more colors a participant already encountered the smaller is her expectation of a new color.

To summarize our findings pertaining to the residual probabilities reported up to this point: (1) drawing a larger sample decreases the residual as more precise information on already observed colors is available; (2) observing more colors, *ceteris paribus*, decreases the residual probability, perhaps because participants feel that the space of so far unobserved events shrinks with the number of observed colors.

In order to better understand to which degree these two factors impact the residual, we estimate a random effects model of the residual on the number of draws and the number of observed colors, controlling for different demographic factors and the

Table 11: Changes of the residuals before vs. after observing a new color.

		Increased	Decreased	Constant	p-value	p-value (corr)
Task 1	\hat{p}_x^2	9	144	21	0.000	0.000
	\hat{p}_x^3	16	56	13	0.000	0.000
	\hat{p}_x^4	25	38	21	0.130	1.000
Task 2	\hat{p}_x^2	3	149	22	0.000	0.000
	\hat{p}_x^3	26	84	59	0.000	0.000
Task 3	\hat{p}_x^2	3	149	22	0.000	0.000
Task 4	\hat{p}_x^2	6	143	25	0.000	0.000
	\hat{p}_x^3	27	94	53	0.000	0.000
	\hat{p}_x^4	27	57	80	0.001	0.013

Matched pairs sign test, p-values corrected by Bonferroni-Holm procedure.

task. Table 12 presents the results of this analysis. Both factors independently have a negative and significant impact on the size of the residual. The impact of the number of outcomes is roughly 11 times stronger than the effect of the sample draws. It is interesting to note that a higher cognitive ability leads to a smaller residual, in contradiction to Hypothesis 6. However, since it is not obvious what “optimal” residual an individual participant should have in this experiment, it is not possible to assess whether it is reasonable to observe this relationship. Finally, we run two robustness checks of the estimations presented in Table 12. First, the results pertaining to the coefficients and their significance remain robust when a fixed-effects instead of a random-effects model is used (Table A6 in Appendix E). Second, conducting the random effects panel analysis solely upon the latter half of the sampling process (only for observations after sampling round 15), leads to a smaller yet still significant negative coefficient for the number of draws (Table A7 in Appendix E). The coefficient for the number of colors observed becomes insignificant, possibly due to lower power with a relatively small number of new colors being observed in later rounds of the sampling process. Taken together, these results indicate that even in the later stages of the sampling process, participants on average reduce their residuals with every additional draw.

Exp.2 - Result 5. *The participants consistently update their residual probabilities downwards, thus, rejecting Hypothesis 3.*

Exp.2 - Result 6. *The participants of higher cognitive ability report smaller residual probabilities, thus, rejecting Hypothesis 6.*

Table 12: Random Effects Estimator: relation between the sample draws and residuals, panel GLS

Size of the residuals	
Num. draws	-0.746** (0.048)
Num. colours observed	-8.402** (0.537)
Cognitive ability	-2.624** (0.547)
Four colours first	-0.914 (2.509)
Constant	77.907** (11.212)
Observations	19,800
Subjects	165

* $p < 0.05$; ** $p < 0.01$

Standard errors in parentheses, standard errors are clustered at the individual level.

Note: The estimation additionally controls for age, gender, being an economics student and risk aversion but the coefficients are not reported.

4.2.3 Bayesian and Reverse Bayesian Rationality

As shown in Subsection 4.2.1, the participants' behavior is consistent with reverse Bayesian reasoning. Does this mean that our participants are in general very rational Bayesian updaters? In order to test this, we study how observing a new color affects the sum of the new residual probability and the estimate of the new color. Technically, before actually observing a new color, its estimate should be included in the estimate of the event *any other color*. Observing a new color can be viewed as unpacking the estimate of the likelihood of yet unobserved colors into two new estimates, an estimate for the new color and another estimate for the yet unobserved colors. That is, in the absence of updating about the joint event, the sum of the estimate of the new color and the new residual should equal the previous residual. Tversky and Koehler (1994) and Sonnemann et al. (2013) find that the sum of such two unpacked estimates violates this principle in a context without learning; the sum of the unpacked estimates often exceeds the 'packed' estimate.

We study such unpacking effects in our data. We define \hat{p}_{Sum}^u as the sum of the new residual and the estimate for the newly observed color C , i.e. $\hat{p}_{Sum}^u = \hat{p}_x^u + \hat{p}_C^u$. As discussed, \hat{p}_{Sum}^u should equal the previous-round residual \hat{p}_x^o if there is no learning. In general, early updates should imply larger learning effects. Thus, the learning component of \hat{p}_{Sum}^u should decrease with the number of draws conducted. Table 13 tests if the ratios $\frac{\hat{p}_{Sum}^u}{\hat{p}_x^o}$ are different from 1. This is strongly the case for all instances of the updates. There is also a significant positive correlation between the number of colors observed and the unpacking ratio (Kendall's $\tau_A = 0.667, \tau_B = 0.785, p = 0.0095$). That is, the unpacking effects get more pronounced as more colors are observed, suggesting a substantial psychological unpacking effect in the spirit of Tversky and Koehler (1994), rather than a rational learning effect. Figure 5 illustrates the size of the effect for the color observed fourth. The estimate of the new color is virtually added to the previous estimate of the event *any other color*, keeping the latter estimate close to constant.²¹ This suggests that our participants are indeed prone to violations of rational updating principles in the current context. As we have argued above, several factors may lead participants to violate the reverse Bayesian principles. This makes the strong evidence for reverse Bayesianism all the more remarkable.

Exp.2 - Result 7. *The participants succumb to the ‘unpacking’ violation of rationality. Thus, behavior consistent with reverse Bayesianism is not a part of uniform adherence to the principles of Bayesian updating.*

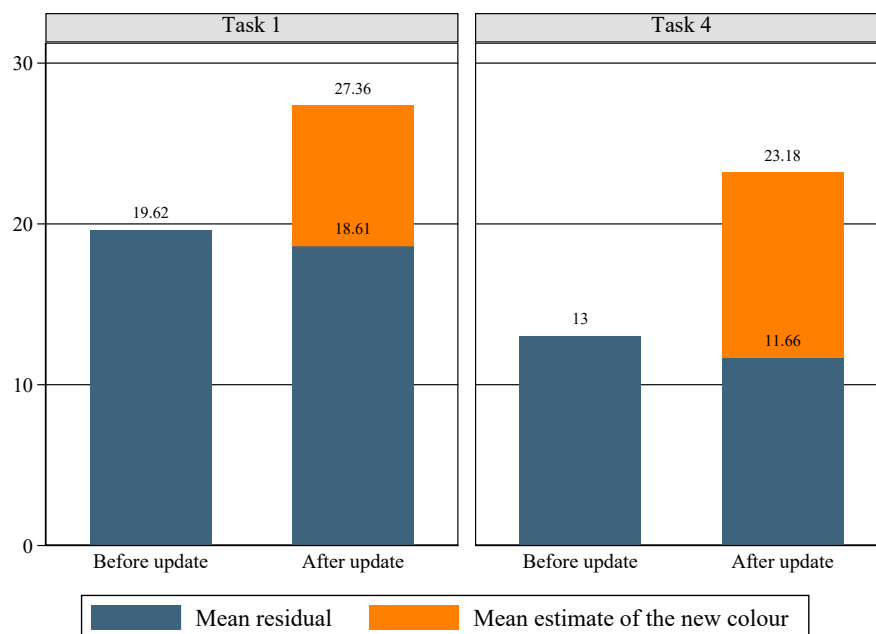
Table 13: Unpacking the residual after observing a new color.

	Second colour	Third colour	Fourth colour
Task 1	1.23	1.46	2.32
Task 2	1.29	1.65	
Task 3	1.48		
Task 4	1.25	2.02	2.23

Wilcoxon signed-rank test, p-values corrected by Bonferroni-Holm procedure, all $p < 0.001$.

²¹See Figures A6 and A7 in Appendix E for an illustration of this when the second and third colors are observed.

Figure 5: Residuals and estimates of a new color before and after observing a new color, fourth color.



5 Concluding Remarks

There is a large literature assessing how decision makers update beliefs about known events in empirical and experimental decision situations (e.g., Charness and Levin, 2005; Charness et al., 2007; Grether, 1992; Holt, 2009). We focus on new and more or less unforeseeable events. Different strands of the theoretical literature offer varying prescriptions on how to integrate information about unforeseen events into beliefs. In their seminal work, Karni and Vierø (2013; 2017) axiomatize a preference functional of a decision maker who integrates unforeseen events into an updated probability distribution over an expanded state space so that the ratio of previously observed events stays unchanged. Our results provide evidence that reverse Bayesianism is also compelling from a descriptive perspective. This stands in sharp contrast to many studies on Bayesian updating, which often find behavior in the lab violating theoretical prescriptions (Charness and Levin, 2005; Charness et al., 2007; Holt, 2009). In other words, our results suggest that reverse Bayesianism is compelling, both, *normatively* and *descriptively*. This holds true both in situations involving reasonably unforeseeable events (Experiment 1) and in situations with unknown but foreseeable

events (Experiment 2).

These implications are intriguing. The space of possible events is continuously expanding in many decision environments. Examples like the financial crisis of 2007 and the COVID-19 pandemic highlight the difficulty to react to novel events appropriately. It is thus important to pay special attention to the possibility of not knowing relevant events beforehand. While our experiments can not speak to the optimal reactions towards an objectively unforeseeable event, they show that decision makers have the capacity to reconcile new information optimally with their existing beliefs: encountering unforeseen events does not lead to a rearrangement of the “old” world of previously known events. Notably, this is irrespective of whether participants started with not explicitly expecting a surprise (Experiment 1) or if they were asked about such a belief and supplied a positive estimate (Experiment 2). Furthermore, it is noteworthy that the participants in Experiment 2 reduced their residual as more draws were made. This indicates an inclination towards becoming more complacent during the experiment, as participants lower their expectations of surprises as fewer new outcomes are discovered. Testing this tendency in more applied scenarios and studying how it interacts with precautionary measures could be an interesting extension to our findings.

Our findings for Experiment 1 also indicate that participants exhibit a higher *WTA* when they are aware of the possibility of further, unknown surprises. In the interpretation of Viscusi and Chesson (1999), the respondents in our study seem to be more hopeful rather than fearful towards unknown future events. This is also interesting in light of the ambiguity literature (Trautmann and van de Kuilen, 2015). One interpretation is that unforeseen events are assigned small probabilities (see also Experiment 2), and that the observed behavior is an embodiment of ambiguity seeking. Indeed, the Trautmann and van de Kuilen (2015) review reports predominantly ambiguity seeking behavior for low probability gain prospects like the ones used in Experiment 1.

In Experiment 2 (and implicitly in Experiment 1) we used the residual probability estimate as a catch-all way of encoding the participants’ beliefs about all remaining possible events. This does not, however, give us a clear description of what exactly participants expect in the future, instead it is a “(...) Black Box, a residual of unknown content.” (Shackle, 1992, p. 23). For example, a participant revealing a positive residual could be expecting a surprise in the future. Say she could be expecting the

urn to also contain blue and red marbles. Observing a purple marble could present her with an unforeseen event. In our design, we do not elicit the precise nature of the unforeseen event - whether it corresponds to purple or some other specific color. Further studies could try to elicit an exhaustive list of expected events from the participants or even try to use a completely non-distributional approach to assess uncertainty (Shackle, 1992).

Furthermore, the surprises in our experiments might still be considered to be easily comprehensible. A new urn with new prizes and a new color in a box of colored marbles can be surprising and unexpected, but it is relatively simple to integrate their materialization into an existing belief structure after their first occurrence. A natural next step could be to study whether the principle of reverse Bayesianism extends to more complex events, where it is less clear which form a surprise might take and, importantly, how to reconcile it with existing beliefs. This could help us to better understand how belief systems are affected by very rare and novel events.

Finally, our results cannot resolve the question whether participants act as-if they are reverse Bayesian or if their underlying thought process adheres to the prescriptions of reverse Bayesianism. On the one hand, some participants did not alter their estimates at all after observing a new outcome. This could suggest the as-if interpretation. On the other hand, a significant share of the participants did, and still provided belief ratios that were either constant or deviated little from the previous ratio. This would be more consistent with decision makers consciously following reverse Bayesianism. More research is needed to further disentangle these two possibilities.

References

- Alaoui, L. and Penta, A. (2016). Endogenous depth of reasoning, *The Review of Economic Studies* **83**(4): 1297–1333.
- Bayarri, M., Benjamin, D. J., Berger, J. O. and Sellke, T. M. (2016). Rejection odds and rejection ratios: A proposal for statistical practice in testing hypotheses, *Journal of Mathematical Psychology* **72**: 90–103.
- Becker, G. M., DeGroot, M. H. and Marschak, J. (1964). Measuring utility by a single-response sequential method., *Behavioral Science* **9**(3): 226–232.
- Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases, *Handbook of Behavioral Economics: Applications and Foundations 1* **2**: 69–186.
- Bors, D. A. and Stokes, T. L. (1998). Raven’s advanced progressive matrices: Norms for first-year university students and the development of a short form, *Educational and Psychological Measurement* **58**(3): 382–398.
- Chambers, C. P. and Hayashi, T. (2018). Reverse Bayesianism: A comment, *American Economic Journal: Microeconomics* **10**(1): 315–24.
- Charness, G., Karni, E. and Levin, D. (2007). Individual and group decision making under risk: An experimental study of bayesian updating and violations of first-order stochastic dominance, *Journal of Risk and uncertainty* **35**(2): 129–148.
- Charness, G. and Levin, D. (2005). How psychological framing affects economic market prices in the lab and field, *The American Economic Review* **95**(4): 1300–1309.
- Chen, D. L., Schonger, M. and Wickens, C. (2016). otree - an open-source platform for laboratory, online and field experiments, *Journal of Behavioral and Experimental Finance* **9**: 88–97.
- Dekel, E., Lipman, B. L. and Rustichini, A. (1998). Standard state-space models preclude unawareness, *Econometrica* **66**(1): 159–173.
- Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on, *Perspectives on Psychological Science* **6**: 274–290.

-
- Dietrich, F. (2018). Savage's theorem under changing awareness, *Journal of Economic Theory* **176**: 1–54.
- Dominiak, A. and Tserenjigmid, G. (2021). Ambiguity under growing awareness, *Journal of Economic Theory* p. 105256.
- Eckel, C. C. and Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices., *Journal of Economic Behavior & Organization* **68**(1): 1–17.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments, *Experimental economics* **10**(2): 171–178.
- Gennaioli, N. and Shleifer, A. (2018). *A Crisis of Beliefs: Investor Psychology and Financial Fragility.*, Princeton University Press.
- Gill, D. and Prowse, V. (2016). Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis, *Journal of Political Economy* **124**(6): 1619–1676.
- Grant, S., Meneghel, I. and Tourky, R. (2017). Learning under unawareness, *Available at SSRN 3113983* .
- Grant, S. and Quiggin, J. (2013). Inductive reasoning about unawareness, *Economic Theory* **54**(3): 717–755.
- Grant, S. and Quiggin, J. (2015). A preference model for choice subject to surprise, *Theory and Decision* **79**(2): 167–180.
- Greiner, B. (2015). An online recruitment system for economic experiments., *Journal of the Economic Science Association* **1**(1): 114–125.
- Grether, D. M. (1992). Testing bayes rule and the representativeness heuristic: Some experimental evidence, *Journal of Economic Behavior & Organization* **17**(1): 31–57.
- Halpern, J. Y. and Rêgo, L. C. (2008). Interactive unawareness revisited, *Games and Economic Behavior* **62**(1): 232–262.
- Heifetz, A., Meier, M. and Schipper, B. C. (2006). Interactive unawareness, *Journal of economic theory* **130**(1): 78–94.

-
- Hoffrage, Ulrich; Hertwig, R. and Gigerenzer, G. (2000). Hindsight bias: A by-product of knowledge, *Journal of Experimental Psychology* **26**(3): 556–581.
- Holt, Charles A. and Smith, A. M. (2009). An update on bayesian updating, *Journal of Economic Behavior & Organization* **69**(2): 125–134.
- Isoni, A., Graham, L. and Sugden, R. (2011). The willingness to pay—willingness to accept gap, the” endowment effect,” subject misconceptions, and experimental procedures for eliciting valuations: Comment., *American Economic Review* **101**(2): 991–1011.
- Karni, E. (2009). A mechanism for eliciting probabilities, *Econometrica* **77**(2): 603–606.
- Karni, E., Valenzuela-Stookey, Q. and Vierø, M.-L. (2020). Reverse bayesianism: A generalization., *The B.E. Journal of Theoretical Economics* **forthcoming**.
- Karni, E. and Vierø, M.-L. (2013). “Reverse Bayesianism”: A choice-based theory of growing awareness, *American Economic Review* **103**(7): 2790–2810.
- Karni, E. and Vierø, M.-L. (2015). Probabilistic sophistication and reverse bayesianism, *Journal of Risk and Uncertainty* **50**(3): 189–208.
- Karni, E. and Vierø, M.-L. (2017). Awareness of unawareness: a theory of decision making in the face of ignorance, *Journal of Economic Theory* **168**: 301–328.
- Kochov, A. (2010). A model of limited foresight, *Technical report*, working paper, University of Rochester.
- Modica, S. and Rustichini, A. (1999). Unawareness and partitional information structures, *Games and Economic behavior* **27**(2): 265–298.
- Ortoleva, P. (2012). Modeling the change of paradigm: Non-bayesian reactions to unexpected news, *American Economic Review* **102**(6): 2410–36.
- Piermont, E. (2021). Unforeseen evidence, *Journal of Economic Theory* (193): 105235.
- Raven, J., Raven, J. C. and Court, J. (1998a). *Manual for Raven’s Progressive Matrices and Vocabulary Scales. Section 4: The Advanced Progressive Matrices.*, Oxford, UK: Oxford Psychologists Press; San Antonio, TX: The Psychological Corporation.

-
- Raven, J., Raven, J. C. and Court, J. (1998b). *Manual for Raven's progressive matrices and vocabulary scales. Section 1: General overview.*, Oxford, UK: Oxford Psychologists Press; San Antonio, TX: The Psychological Corporation.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D. and Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis, *Psychonomic bulletin & review* **16**(2): 225–237.
- Schipper, B. C. (2013). Awareness-dependent subjective expected utility, *International Journal of Game Theory* **42**(3): 725–753.
- Schipper, B. C. (2022). Predicting the unpredictable under subjective expected utility, *Mimeo* .
- Shackle, G. L. S. (1949). *Expectation in economics.*, Cambridge University Press.
- Shackle, G. L. S. (1992). *Epistemics and economics: A critique of economic doctrines.*, Transation Publishers.
- Sonnemann, U., Camerer, C. F., Fox, C. R. and Langer, T. (2013). How psychological framing affects economic market prices in the lab and field, *PNAS* **110**(29): 11779–11784.
- Trautmann, S. T. and Schmidt, U. (2012). Pricing risk and ambiguity: The effect of perspective taking., *Quarterly Journal of Experimental Psychology* **65**(1): 195–205.
- Trautmann, S. T. and van de Kuilen, G. (2015). Ambiguity attitudes., in G. Keren and G. Wu (eds), *The Wiley Blackwell Handbook of Judgment and Decision Making*, Blackwall, chapter 3, pp. 89–116.
- Trautmann, S. T., Vieider, F. M. and Wakker, P. P. (2011). Preference reversals for ambiguity aversion., *Management Science* **57**(7): 1320–1333.
- Tversky, A. and Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability, *Psychological Review* **101**(4): 547–567.
- Viscusi, W. K. and Chesson, H. (1999). Hopes and fears: the conflicting effects of risk ambiguity., *Theory and decision* **47**(2): 157–184.

Reverse Bayesianism: Revising Beliefs in Light of Unforeseen Events

Online Appendix

(Not meant to be part of publication)

Christoph K. Becker, Tigran Melkonyan, Eugenio Proto, Andis Sofianos, Stefan T. Trautmann

May 7, 2022

A	Instructions for Experiment 1	A.2
B	Instructions for Experiment 2	A.10
C	Example of sampling screen in Experiment 2	A.16
D	Additional analysis for Experiment 1	A.17
E	Additional analysis for Experiment 2	A.22

A Instructions for Experiment 1

Note: In *red* are comments, text highlighted in *yellow* relates to only the IS condition and text highlighted in *green* relates to only the PS condition

General Information

Welcome to the experiment.

Thank you for volunteering your time to participate in this experimental project. The purpose of this experiment is to study how people make decisions in a particular situation. The results of this experiment will have applications to behavioral economics and economics in general.

During this experiment, please follow the instructions very carefully. Please remain silent during the session. You will go through various stages where in some instances you will simply observe outcomes of draws from an urn and in other cases you will be asked to report your perceived likelihood of an event or how much an item is worth to you. These reports, which you will enter on your computer, will determine your eventual monetary reward from participating in this experiment.

You may have heard about experiments in which participants were deceived. This experiment **does not** involve deception by the experimenters. That is, everything the experimenter tells you, and all on-screen instructions, are true and accurate.

Initial Instructions

It is critical that you read through these instructions carefully as fully understanding them will allow you to substantially increase your eventual monetary payoff from this study, where you can earn from a minimum of €4.00 to a maximum of €26.00 depending on your decisions.

During this study you will be asked to observe some outcomes of draws from an urn and will then be asked to report how likely certain 'events' are and how much some 'items' are worth to you. One of these choices will be randomly picked to determine your monetary payment for completing this study. Depending on your responses you stand to earn a substantial amount of money. Over the next few pages we explain how your earnings will be determined. Please read this very carefully.

Likelihoods of events – Reporting and Earnings

At different instances during this study, you will be asked to provide us with your perceived likelihood of an outcome of a random draw from an urn. The urn will contain prizes of varying monetary value. You will have an opportunity to observe multiple random draws out of this urn to gain a good understanding of the likelihood of different prizes.

After observing a sequence of random draws from the urn, you will be asked to report your beliefs about the likelihood that the prize drawn from the urn is of particular value. For example, we may ask you to report your perception of the likelihood that the prize drawn out of the urn has a value of €30.

You will be asked to report a number between 0 and 1. The closer to the true value your reported number is the greater would be your expected potential bonus from this decision.

Your best strategy is to estimate your own perceived likelihood and truthfully report that likelihood.

Suppose that we ask you to report the likelihood of drawing a ball with prize X. Your reported likelihood will be compared to a likelihood randomly generated by the computer. This **randomly generated likelihood** will be a number between 0 and 1 and it will be completely unrelated to your reported likelihood. If your reported likelihood is greater than or the same as the randomly generated likelihood, you will be endowed with a lottery that pays you X with a probability equal to the actual proportion of prize X in the urn and pay you nothing otherwise. If, instead, your reported likelihood is lower than the randomly generated likelihood, you will be endowed with a lottery that pays you X with a probability equal to the randomly generated likelihood and pay you nothing otherwise. After these choices are made and revealed by the computer, the lottery in your (virtual) possession will be played and payments made according to the realization of the lottery.

Example: Let's say we ask you to provide your perceived likelihood that a prize of value €30 will be randomly drawn from the urn. Let's further suppose that the true proportion of prize €30 is 0.5 in the urn. If your reported likelihood is 0.4 and the randomly generated likelihood is 0.3, you will receive the lottery that pays you €30 with probability 0.5 (true proportion of the prize) and pay you nothing otherwise. If your reported likelihood is 0.4 and the randomly generated likelihood is 0.6, you will receive the lottery that pays you €30 with probability 0.6 (the randomly generated likelihood) and pay you nothing otherwise.

Values of 'items' – Reporting and Earnings

At different instances during this study, you will be asked to provide us with the value at which you would be willing to sell an 'item'. This 'item' will be a lottery that would give you some monetary prizes with some probabilities. During the tasks that will follow, the prizes that will be possible to earn and their likelihoods will not be explicitly stated and so you would need to rely on what you expect.

For example, if you believe that there is equal likelihood of two prizes of value of €20 or €10, then the corresponding lottery would entail a payoff of €20 with probability 0.5 or a payoff of €10 with probability 0.5.

You will be given the lottery and it will be your task to provide us with a value which you would feel comfortable to sell us back this lottery. Your decision is essentially to provide us with the certain amount that you would be happy to receive instead of playing the lottery at the particular instance.

As you will see, your best strategy is to provide us with the minimum amount you would be willing to receive for selling us the lottery

Your named amount will be compared to a fixed amount. This fixed amount will be randomly generated by a computer and will be completely unrelated to your named amount:

- If your named amount is less than or the same as the fixed amount, then you get to sell the lottery. But, here's the interesting part. You do not receive the amount you offered. Instead, we pay you the fixed amount, i.e. the randomly generated amount which is higher than or equal to your offer.
- If, on the contrary, your named amount is more than the fixed amount then you don't get to sell the lottery and will be paid according to the realization of the lottery.

Example: if your named amount is €50 and the fixed amount is €60, you get to sell the lottery and receive the certain amount of €60. if your named amount is €50 and the fixed amount is €40, you do not get to sell the lottery and thus receive payment according to the realization of the lottery.

You should offer the minimum amount you would be willing to accept in exchange for the lottery you own. Your best strategy is to determine your personal value for the item and record that value as your offer. **It will not be to your advantage to suggest more than this amount, and it will not be to your advantage to suggest less.** There is not necessarily a “correct” value. Personal values can differ from individual to individual.

Example of best strategy for deciding valuation

The following example illustrates how you work out the minimum you are willing to accept for a lottery.

Imagine that I am a seller of a lottery “A”. How do I know the minimum I’d be willing to sell lottery “A” for?

Start with 1 penny. Would I be willing to get 1 penny for the item? If NOT, then increase the amount to 2 pence. If I’m NOT willing to accept 2 pence, then increase further. I keep increasing until I come to an amount that makes me indifferent between keeping lottery “A” or getting a certain amount.

Example: Would I sell lottery “A” for €1.00? NO. So I need to consider higher amounts. Would I sell lottery “A” for €2.00? YES. Would I sell lottery “A” for €1.90? YES, Would I sell lottery “A” for €1.80, YES. Would I sell lottery “A” for €1.50? I don’t care whether I end up with €1.50 or keep the lottery. Then that is the minimum I’d be willing to accept for lottery “A”. I’ll record that number on the computer.

The key to determining the minimum you’d be willing to accept is remembering that you will not necessarily get only the amount you declare. Instead, if you receive anything, you will receive the fixed offer.

Why is my best strategy to declare the minimum I’d be willing to accept? Let’s go back to the example:

Say that I decide that the minimum I’d be willing to accept for lottery “A” is €1.50.

What happens if I declare more than €1.50? Say I declare €2.

If the fixed amount is, say, €1.90, then I don’t sell the lottery. Had I declared €1.50, I would have received the amount €1.90 for a lottery that I think is worth €1.50. So I lose out.

What happens if I declare less than €1.50? Say I bid €1.00.

If the fixed offer is €1.20, then I have to accept €1.20 for a lottery that I really think is worth €1.50. I lose out.

Payment procedures

You will be asked to provide a value for lotteries and likelihood for events at different instances as we described in the previous pages. One out of all these decisions will be randomly chosen and payments will be made according to that decision and the realisation of the relevant lottery.

All prizes and lotteries that you will be asked to consider and make decisions about will be expressed in tokens. Each token corresponds to €0.05. Thus, a prize of 100 tokens will be equivalent to €5.

PART 1

You will now simply observe random draws of balls out of an urn. This urn contains a number of balls. Each ball represents a potential prize in terms of payment in tokens. You will observe 20 consecutive random draws with replacement from the urn. Please pay attention to the prizes that will appear and their frequencies. **For IS conditions:** This urn contains two and only two possible prizes. Your earnings do not directly depend on the outcome of each random draw in this stage, but understanding the composition of the urn may considerably improve your future earnings. **For PS conditions:** At any point in the study new balls representing different tokens to what you have been observing so far may be added to this urn. Please click OK when you are ready to proceed.

Belief Screen 1

We would now like to ask you some questions about the likelihoods of different prizes in the urn you have been observing.

Remember that your most profitable strategy is reporting truthfully your assessment of different likelihoods.

You can remind yourself of the payment procedure and instructions related to the task by referring back to the instructions in front of you.

What is your estimate of the likelihood that prize 80 is drawn from the urn? _____

What is your estimate of the likelihood that prize 190 is drawn from the urn? _____

WTA Screen 1

We would like to ask for your value of a lottery.

Recall that when answering questions about your value of a lottery, your named amount will be compared to a fixed amount. If your named amount is greater than or equal to the fixed amount then you do not sell the lottery and are thus paid according to the realization of the lottery. Otherwise, if your named amount is less than the fixed amount, you get to sell the lottery you have been endowed with and receive the fixed amount as a payment.

Remember that your most profitable strategy is reporting truthfully your valuations, you can remind yourself of the payment procedure and instructions related to the task by referring back to the instructions in front of you.

Thinking about the different prizes and the composition of the urn you have just observed.

What is the minimum amount you are willing to accept to sell the lottery that pays according to a draw from the urn? _____

After giving a choice, the following appears on the screen

Thank you. Your choice has been recorded.

For IS conditions:

We will now draw another prize from the urn.

If the decision you just made is selected by the computer to be the payment relevant round, the draw about to take place will be used to determine your payment.

You will not be shown the prize drawn out of the urn in this instance. Your colleague making the draws will make a note of the prize drawn to be used later if necessary.

We make a draw out of the original urn and the participant making the draws notes down the prize drawn.

At this point we bring the new urn, draw one prize and say:

On PC Screen:

*This urn contains **only** the prize you are about to be shown. Please click OK to confirm you understand this.*

Now we drop the contents of the original urn into the new urn which together form the updated urn.

For PS conditions:

If the decision you just made is selected by the computer to be the payment relevant round, the draw about to take place from the urn in the front will be used to determine your payment.

At this point we bring the new urn, draw one prize and say:

On PC Screen:

This urn contains new prizes. The urn contains no prizes similar to what you have been observing as a result of random draws from the other urn. Please click OK to confirm you understand this.

Now we drop the contents of the original urn into the new urn which together form the updated urn.

We will now draw a prize from the urn.

If the decision you just made is selected by the computer to be the payment relevant round, this draw will be used to determine your payment. You will not be shown the prize drawn out of the urn in this instance. Your colleague making the draws will make a note of the prize drawn to be used later if necessary.

A random draw from the updated urn is made and the participant making the draws notes down the prize drawn.

Belief Screen 2

We would now like to ask again some questions about the likelihoods of different prizes in the urn you have been observing.

Remember that your most profitable strategy is reporting truthfully your assessment of different likelihoods.

You can remind yourself of the payment procedure and instructions related to the task by referring back to the instructions in front of you.

What is your estimate of the likelihood that prize 80 is drawn from the urn? _____

What is your estimate of the likelihood that prize 190 is drawn from the urn? _____

What is your estimate of the likelihood that prize 15/375 is drawn from the urn? _____

WTA Screen 2

Remember that your most profitable strategy is reporting truthfully your valuations.

You can remind yourself of the payment procedure and instructions related to the task by referring back to the instructions in front of you.

Again thinking about the urn in front of you.

What is the minimum amount you are willing to accept to sell the lottery that pays according to a random draw from the urn? _____

After giving a choice, this appears:

Thank you. Your choice has been recorded

We will now draw another prize from the urn.

If the decision you just made is selected by the computer to be the payment relevant round, this draw will be used to determine your payment. You will not be shown the prize drawn out of the urn in this instance. Your colleague making the draws will make a note of the prize drawn to be used later if necessary.

A random draw from the updated urn is made and the participant making the draws notes down the prize drawn.

PART 2

Lottery Choice Task

On your screen below you see a list of 6 lotteries with the prizes given in terms of tokens. You have to make a choice among these 6 lotteries. For each of the listed below lotteries, the chance for either of the two payoffs is equal. That is, for lottery 2 for example, you can win 24 tokens with 50% chance and 36 tokens with 50% chance. Your chosen lottery will be played out and you will be paid according to the realization of that lottery. As before, each token corresponds to €0.05. Thus, for a prize of 100 tokens the equivalent dollar amount will be €5.

<u>Lottery</u>	<u>X</u>	<u>Y</u>
1	28	28
2	24	36
3	20	44
4	16	52
5	12	60
6	2	70

PART 3

Short Raven Test implemented

PART 4

General Demographic Questionnaire

- How old are you? (years)
- What is your gender? (M/F/Other [Please describe if you wish]/Prefer not to disclose)
- What is your country of origin?
- What is your religion?
 - Buddhist
 - Christian
 - Hindu
 - Jewish
 - Muslim
 - Sikh
 - No religion
 - Other [Please describe if you wish]
 - Prefer not to disclose
- What is your field of studies/major?
- What is your year of study?
- In high school, what was the highest possible grade? (E.g. A, 100, 20)
- What was your final grade in high school?
- In political matters, people talk of “the left” and “the right”. How would you place your views on this scale, generally speaking?

Table 14: Sample draws per session & Average WTAs Reported

Session	Treatment	Draw 1	Draw 2	Draw 3	Draw 4	Draw 5	Draw 6	Draw 7	Draw 8	Draw 9	Draw 10	Draw 11	Draw 12	Draw 13	Draw 14	Draw 15	Draw 16	Draw 17	Draw 18	Draw 19	Draw 20	WTA ^a	WTA ^w
1	IS, low prize	190	190	190	190	190	190	80	80	80	190	190	190	190	190	190	190	190	190	80	80	188.75	157.5
2	IS, low prize	80	80	190	190	190	190	190	190	80	80	190	190	190	190	190	190	190	190	80	80	94.875	72.625
3	IS, low prize	80	190	80	80	80	80	80	80	80	190	190	80	190	190	190	80	190	190	80	190	108.105	731.053
4	IS, low prize	80	190	190	80	80	80	190	80	190	80	190	190	190	190	190	190	190	190	190	190	119.364	756.364
5	IS, low prize	190	190	190	80	80	80	80	190	80	190	80	190	190	190	80	190	80	190	80	190	871.538	858.462
6	IS, low prize	190	80	190	80	190	80	190	80	190	80	190	190	190	190	80	190	80	80	80	80	134.429	125
7	IS, low prize	190	80	80	190	80	190	190	190	190	80	190	190	190	190	190	190	80	80	80	80	112.667	943.333
8	IS, high prize	80	80	80	190	190	80	190	80	190	190	190	190	190	190	190	190	80	80	190	80	115.333	164.667
9	IS, high prize	190	190	190	190	190	80	80	80	190	190	80	190	80	190	80	190	80	190	80	190	97.6	161
10	IS, high prize	190	190	190	190	190	190	190	80	190	80	190	80	190	190	190	80	80	80	190	190	752.143	974.286
11	IS, high prize	190	80	190	190	190	190	80	190	80	190	80	190	190	190	190	190	190	190	80	190	152.857	143
12	IS, high prize	190	80	190	80	190	190	190	190	190	190	80	80	190	80	190	80	190	80	190	190	117.714	185.643
13	IS, high prize	80	190	80	80	190	80	190	190	80	190	190	190	80	190	190	190	80	190	80	190	129.8	138
14	IS, high prize	190	80	190	80	190	190	80	190	80	80	190	80	190	80	190	190	190	80	80	80	110.778	171.111
15	PS, low prize	80	80	190	190	80	80	80	190	190	80	190	80	190	80	190	80	190	80	190	190	135	77.5
16	PS, low prize	190	190	190	190	190	190	80	80	190	80	190	80	190	80	190	80	80	80	80	80	124.25	96.05
17	PS, low prize	80	190	190	190	80	190	80	80	190	190	190	190	190	190	190	80	190	190	190	190	127.5	80
18	PS, low prize	190	190	190	80	190	190	80	80	190	80	80	190	190	190	80	80	190	190	80	190	115	766.667
19	PS, low prize	190	80	190	190	190	190	80	80	80	190	190	190	190	80	80	80	190	80	80	80	150	92.5
20	PS, low prize	190	190	190	190	80	190	190	80	190	190	190	190	190	190	190	190	190	80	190	190	154	125.333
21	PS, low prize	80	190	190	80	190	190	190	190	80	190	80	190	80	190	190	190	190	80	80	80	105.667	796.667
22	PS, low prize	80	190	190	80	80	190	80	80	190	80	190	190	190	190	80	190	80	190	190	80	179.286	960.714
23	PS, low prize	190	80	80	80	190	80	190	190	80	80	190	80	80	190	80	190	80	190	80	80	140.5	116
24	PS, low prize	80	80	190	190	190	80	190	80	190	80	190	190	80	190	80	80	80	80	190	190	127.273	823.636
25	PS, high prize	190	190	190	80	80	80	190	190	190	190	190	190	190	190	190	190	190	80	80	80	121.5	157.5
26	PS, high prize	190	80	80	190	190	190	190	190	80	190	80	80	190	80	190	80	80	190	80	190	131.944	173.889
27	PS, high prize	190	190	190	190	190	190	80	80	80	80	190	190	80	190	190	80	190	190	80	190	122.636	161.364
28	PS, high prize	190	190	190	80	190	80	80	80	190	80	80	80	190	190	190	190	190	80	190	80	124.308	174.462
29	PS, high prize	80	80	80	190	80	190	190	190	190	190	190	190	190	190	190	190	190	190	190	80	151.75	186.625
30	PS, high prize	190	190	190	80	190	190	190	80	190	190	190	190	190	80	80	80	190	80	190	190	149.7	180.1
31	PS, high prize	190	190	190	190	190	190	80	190	80	190	80	190	80	190	190	190	190	80	80	190	162.5	211.667
32	PS, high prize	190	80	190	80	190	80	190	80	80	190	190	190	80	190	190	190	190	80	190	190	134.375	175.625
33	PS, high prize	190	80	80	80	80	190	190	80	190	80	80	80	80	190	80	190	80	190	80	190	116.25	246.25

B Instructions for Experiment 2

Note: In red are comments that were not visible to participants.

General Instructions

Thank you for participating in today's experiment.

If you have any questions during the experiment, please raise your hand. An experimenter will approach your table to answer your question in private.

You may have heard about experiments in which participants were deceived. This experiment does not involve deception by the experimenters. That is, everything the experimenter tells you, and all on-screen instructions, are true and accurate.

The experiment consists of 4 parts. For participating in this experiment you will earn £3 at the end of the experiment. In addition you can earn a bonus of £6 in each of the four parts, depending on your performance in the experiment and chance. After these four parts you will play a pattern game, in which you can earn additionally up to £2.

In the end follows a short demographic questionnaire.

Sampling Boxes

The experiment consists of 4 parts. In each part, you draw a random sample of (virtual) marbles from a (virtual) box containing exactly 100 colored marbles. Initially, you have no information about the contents of each box: you do not know which colors, or how many different colors, are in the box. The four parts and four boxes are independent of each other: different boxes are used for different parts.

In each part, you draw 30 marbles with replacement one after another from the box. You draw a marble by clicking the button "Draw" (or by pressing enter). Once clicked, the computer randomly draws a marble from the box. The result of a draw is shown on-screen with a marble of the color and the name of the color.

The sample draws are conducted with replacement. For example, if you drew a magenta marble (this color is not used in the actual experiment) from a box, this marble is placed back in the box for the next draw, such that the number of marbles of each color in the box stays the same as you sample. All marbles you have sampled (and their colors) are registered at the bottom of the screen.

Your payoff-relevant task

After each draw of a new marble, you will be asked to state your expectation about the contents of the box, that is, about the distribution of colors in the box. The more precise your prediction is, the higher will be your expected payoff from the experiment (details below). After each draw, you will be asked to separately indicate:

- (i) Your expected number of marbles in the box for each color that you have already observed for the box, and
- (ii) Your expected number of marbles of “any other colors” that you have not yet observed for the box, and that may or may not be in the box.

Example

Suppose you drew a **magenta** marble in your first draw and a **teal** marble (this color is also not used in the actual experiment) in the second draw. After the first draw you would be asked to guess how many **magenta** marbles are in the box, and how many marbles of any other color, not yet observed, are in the box. After the second draw you would be asked how many **magenta** marbles are in the box, how many **teal** marbles are in the box, and how many marbles of any other color, not yet observed, are in the box.

As the box contains exactly 100 marbles, your estimates of the number of marbles of the already observed colors and of any other colors you may think are in the box (but not yet observed) must add up to 100. Moreover, if you expect the number of marbles of other colors to be zero, you need to explicitly submit an estimate of zero (that is, not just leaving the entry field open).

After the 30th marble is drawn, you will enter your last prediction for this box. A new button “Continue to the next box” will allow you to continue to the next part, with a new box to sample.

Entering estimates in the program

After each draw, you can enter your estimates by typing them into the entry fields. You can also use the “fill previous estimate” buttons to pre-fill your previous round’s estimates for each color. At any point before making the next draw, you can adjust the current estimates in the entry fields using “+” and “-” buttons next to the entry fields.

Getting paid for good predictions

You may earn a bonus of £6 for each part of the experiment. All of your answers provided for all four parts will affect your chances of receiving the bonus. If you want to maximize your expected earnings from this experiment, it is in your best interest to estimate the number of marbles for each box as accurately as possible, and report them truthfully after each draw. To determine whether you will win a bonus, you will draw a marble either from one of the boxes in the experiment (called *Estimate Box*), or from another, newly constructed one (called *New Box*). Importantly, your reported estimates will influence the construction of this new box.

If you report your estimates accurately and truthfully, this will be best for you in terms of your expected payment from the experiment. Below we will explain the payment procedure, and provide the intuition and an example why it is in your best interest to report your estimates as correctly as possible after each draw. You are invited to review these explanations. Please note that they are not necessary to understand the experiment and can be skipped without any harm if you are not interested. You can request a hard copy of these details at any point of the experiment in case of doubt.

Payment procedure (click to expand):

The below was hidden and only visible if the participants chose to expand the information:

After you finished sampling from all four boxes, for each of the four parts you may earn a bonus of £6 as follows:

Estimate box: The computer randomly selects one of the 30 draw rounds, and then randomly selects one color estimate you made for this round (this is the selected color for this task). This can be an estimate for some color you have observed, or alternatively an estimate for the number of not yet observed colors at some point, that is, “any other color”. Note that all of your estimates have the same chance to be randomly selected.

New box: Next, the computer constructs a new box of 100 marbles that contains only two colors, black or white. Every possible combination of black and white marbles (the number of white marbles = 100 – the number of black marbles) is equally likely.

Next, the computer compares the number of black marbles in the New Box with the estimate you made for the selected color in the experiment (or for “any other color”).

- If your estimate for the selected color is larger than the number of black marbles in the New Box, you will draw one marble from the Estimate Box. If this marble is of the

selected color, you will receive £6. If the marble is not of the selected color, you will receive £0.

- If your estimate is smaller than the number of black marbles in the New Box, you will draw one marble from the New Box. If this marble is black, you will receive £6. If the marble is white, you will receive £0.

Intuition (click to expand):

The below was hidden and only visible if the participants chose to expand the information:

You will have the best chance to win the bonus of £6 for each part, by truthfully reporting your estimate. For example, if you think there are many magenta marbles in the Estimate Box, you will more likely make a draw from this box. This is because in your estimation the number of black marbles in the New Box will most likely be smaller than your estimate of magenta marbles for the Estimate Box.

If you think there are only few magenta marbles in the Estimate Box, you will more likely make a draw from the New Box. This is because the number of black marbles in the New Box will most likely be larger than your estimate of magenta marbles for the Estimate Box.

Thus, as long as you report your estimate for each color in each draw and each box accurately and truthfully, the mechanism makes sure that you get the box with the highest chance of winning the bonus.

Note that your winning chance in the case of making the payoff-relevant draw from the Estimate Box depends only on the true number of marbles of that color in the box. Similarly, in the case of making the payoff-relevant draw from New Box, the chance depends only on the number of black marbles in the box. Your estimate of colors for the boxes in the experiment is only relevant for determining the best boxes for you during the payment procedure. Thus, better estimates give you better chances to win.

Example (click to expand):

The below was hidden and only visible if the participants chose to expand the information:

Example - Part 1

For part 1, the computer selected the round 16 draw. In this round you provided estimates of the number of magenta marbles, teal marbles, and the number of marbles of “any other color”.

The computer further selected magenta as the payoff-relevant color estimate. Suppose your estimate of the number of magenta marbles in box 1 in round 16 was 42 marbles.

Suppose the computer randomly generated a New Box that contained 35 black and 65 white marbles. Because 35 black winning marbles in New Box is less than your estimate of 42 magenta winning marbles in Estimate Box 1, your bonus would be determined by Estimate Box 1. Note that your true chance to win the bonus of £6 would depend on the true number of magenta marbles in box 1. Suppose you drew a teal marble from Estimate Box 1. Your bonus for part 1 would be £0.

Example - Part 2

For part 2 box, the computer selected the round 2 draw. In this round you provided estimates of the number of magenta marbles, and the number of marbles of “any other color”. The computer further selected “any other color” as the payoff-relevant color estimate. Suppose your estimate for the number of “any other color” marbles in box 2 in round 2 was 50 marbles.

Suppose the computer randomly generated another New Box that contained 7 black and 93 white marbles. Because 7 black winning marbles in New Box is less than 50 winning marbles of “any other color” in Estimate Box 2, your bonus would be determined by a draw from Estimate Box 2. Note that your true chance to win the bonus of £6 would depend on the true number of - marbles in box 2 that are not magenta. Suppose you drew a teal marble from box 2. Your bonus for part 2 would be £6.

Example - Part 3

For part 3 box, the computer selected the round 30 draw. In this round you provided estimates for the number of magenta marbles, the number of teal marbles, and the number of marbles of “any other color”. The computer further selected teal as the payoff-relevant color estimate. Suppose your estimate for the number of teal marbles in box 3 in round 30 was 20 marbles.

Suppose the computer randomly generated another New Box that contains 87 black and 13 white marbles. Because 87 black winning marbles in New Box is more than 20 twinning marbles of teal color in Estimate Box 3, your bonus would be determined by New Box 3. Suppose you drew a black marble from box 3. Your bonus for part 3 would be £6.

Example - Part 4

For part 4 box, the computer selected the round 7 draw. In this round you provided estimates for the number of magenta marbles, the number of teal marbles, and the number of marbles of “any other color”. The computer further selected “any other color” as the payoff-relevant color

estimate. Suppose your estimate for the number of “any other color” marbles in box 4 in round 7 was 33 marbles.

The computer randomly generated another New Box that contains 27 black and 73 white marbles. Because 27 black winning marbles in New Box 4 is less than 33 winning marbles of “any other color” in Estimate Box 4, your bonus would be determined by a random draw from Estimate Box 4. Note that your true chance to win the bonus of £6 would depend on the true number of marbles in box 4 that were neither magenta nor teal. Suppose you drew a teal marble from box 4. Your bonus for part 4 would be £0.

Pattern game

You will now play a pattern game, where you are asked to solve some puzzles

On the screen, you will see a set of abstract pictures with one of the pictures missing. You need to choose a picture from the choices below to complete the pattern.

You will have a total of 8 minutes to complete 12 such puzzles.

During these 8 minutes you will be able to move forwards and backwards and change your answers using the buttons and tabs on your screen.

At the end of the experiment, the computer will randomly draw two of the puzzles from the pattern game. Each puzzle has the same probability to be chosen. For each of the two puzzles that you solved correctly, you will earn an additional £1.


Once the 8 minutes have passed, the pattern game will be automatically submitted and you will proceed to the results. You can submit all your answers and wait for the others to finish once you reach the last puzzle by clicking on the button that will appear and be labelled "Finish and go to results".

C Example of sampling screen in Experiment 2

Part 1

Please draw a sample from the box.

Sample draw: 30


maroon

Please indicate in the fields below, how many marbles of a samples color you think are in this box. Remember, the box has a total of 100 marbles.


Number of **orange** marbles: [Fill previous estimate](#) [+](#) [-](#)

Number of **maroon** marbles: [Fill previous estimate](#) [+](#) [-](#)

Number of **blue** marbles: [Fill previous estimate](#) [+](#) [-](#)

Number of **salmon** marbles: [Fill previous estimate](#) [+](#) [-](#)

Number of marbles of **any other color**: [Fill previous estimate](#) [+](#) [-](#)



[Go to part 2](#)

Note: Example of the sampling screen in the experiment, first task, four outcomes first treatment. The example depicts the screen after the 30th sample draw.

D Additional analysis for Experiment 1

Figure A1: Histograms of the residual in the original urn ($\hat{p}_x^o = 1 - \hat{p}_{80}^o - \hat{p}_{190}^o$)

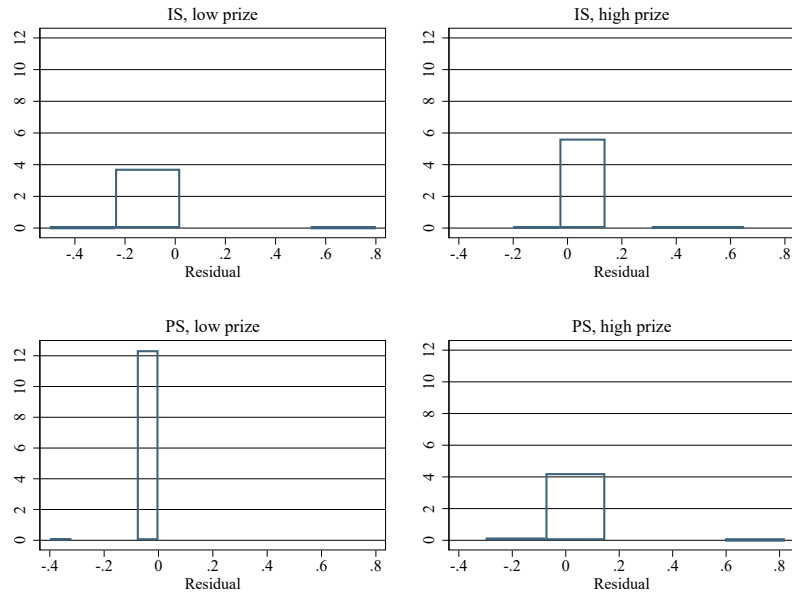


Figure A2: Histograms of the residual in the updated urn ($\hat{p}_x^u = 1 - \hat{p}_{80}^u - \hat{p}_{190}^u - \hat{p}_s^u$)

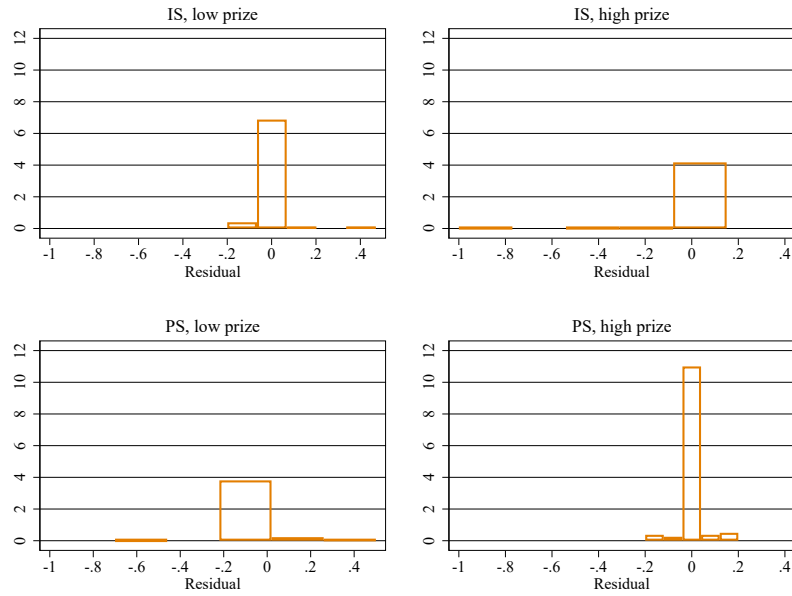


Table A1: Relation between the ratio differences and cognitive ability; OLS regression.

Ratio differences ΔR	
Cognitive ability	-0.008 (0.006)
Constant	0.221* (0.099)
Observations	343

* $p < 0.05$; ** $p < 0.01$

Standard errors in parentheses

Note: The estimation additionally controls for high prize, age, gender, being an economics student and risk aversion but the coefficients are not reported.

Figure A3: Histograms of the change in the ratios before vs. after the urn is updated, by treatment. Histogram in blue, box plot in orange, outliers (circles) and mean (diamond) in black.

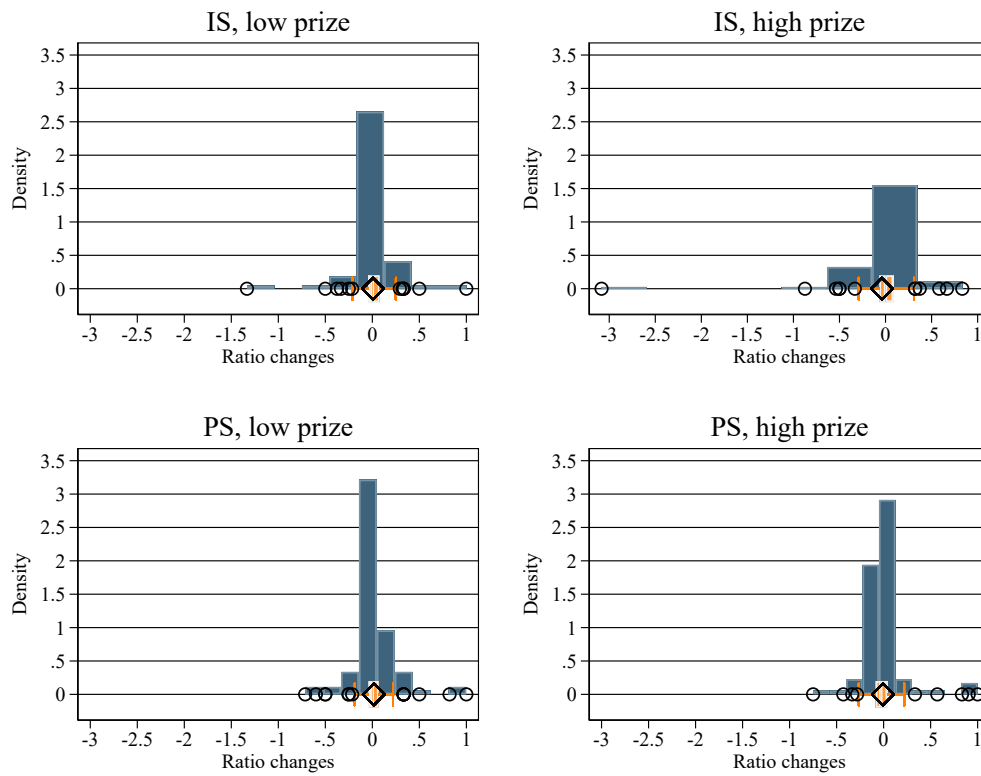


Table A2: Relation between *WTA* and possible moderators, original urn; OLS regression.

<i>WTA</i>	All treatments	Low prize	High prize
PS	23.770** (6.423)	25.511* (10.406)	23.797** (8.052)
# prizes 190 observed	-2.744 (2.304)	-1.238 (3.744)	-5.475 (3.228)
Belief about # prizes 190	107.165** (37.626)	90.308 (68.299)	133.747** (43.275)
Cog Ability	0.800 (1.527)	2.471 (2.619)	-1.416 (1.872)
Age	-1.792* (0.773)	-1.059 (1.127)	-2.869* (1.109)
Female	-20.848** (6.658)	-24.145* (10.985)	-18.878* (8.300)
Econ	1.172 (7.333)	4.272 (11.500)	-1.352 (9.616)
Risk aversion	-1.200 (2.856)	0.967 (4.415)	-4.824 (3.835)
Constant	126.481** (29.600)	87.957* (43.277)	186.385** (44.218)
Observations	344	169	175

* $p < 0.05$; ** $p < 0.01$

Standard errors in parentheses

Table A3: Relation between *WTA* and possible moderators, updated urn; OLS regression.

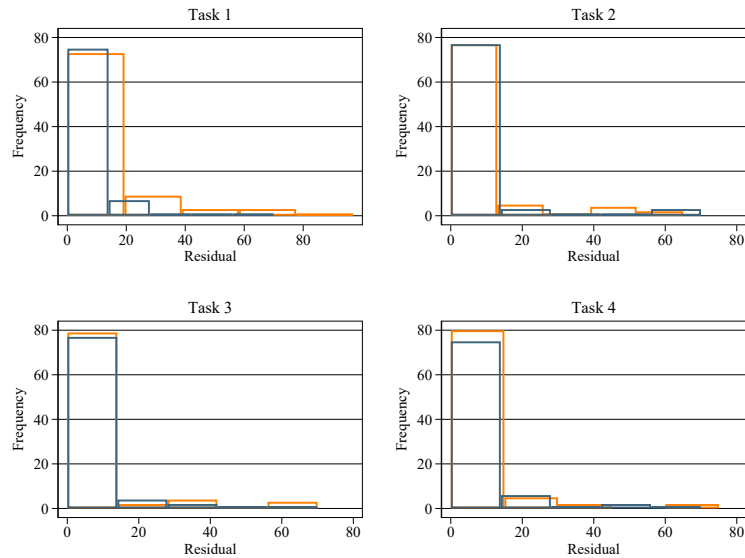
<i>WTA</i>	All treatments	Low prize	High prize
PS	16.509* (7.722)	4.807 (6.924)	32.303** (10.866)
# prizes 190 observed	-0.420 (2.404)	-3.264 (2.046)	-6.300 (3.868)
Belief about # prizes 190	80.578* (36.053)	140.121** (35.405)	-36.993 (47.142)
Cog Ability	2.001 (1.838)	5.250** (1.767)	0.365 (2.486)
Age	-1.110 (0.930)	0.511 (0.756)	-1.665 (1.461)
Female	-10.362 (7.983)	-5.076 (7.374)	-27.371* (11.097)
Econ	0.164 (8.785)	16.616* (7.691)	-2.022 (12.810)
Risk aversion	-7.285* (3.400)	-3.280 (2.879)	-11.742* (5.139)
Constant	110.841** (35.545)	14.018 (28.851)	304.474** (58.634)
Observations	344	169	175

* $p < 0.05$; ** $p < 0.01$

Standard errors in parentheses

E Additional analysis for Experiment 2

Figure A4: Histograms of the residual \hat{p}_x^{30} (after observing the last sample draw).



The orange boxes show the residuals of *two colors* treatment while the the blue boxes show the residuals of *four colors* treatment.

Table A4: Relation between the ratio differences and cognitive ability, panel GLS.

Ratio differences ΔR	
Cognitive ability	0.028 (0.026)
Num. draws	0.022* (0.010)
Num. colours observed	-0.066 (0.123)
Constant	-0.314 (0.651)
Observations	1,119
Subjects	

* $p < 0.05$; ** $p < 0.01$

Standard errors in parentheses

Note: The estimation additionally controls for age, gender, being an economics student and risk aversion but the coefficients are not reported.

Figure A5: Histograms of the changes in the ratios before vs. after the urn is updated, by treatment. Histogram in blue, box plot in orange, outliers (circles) and mean (diamond) in black.

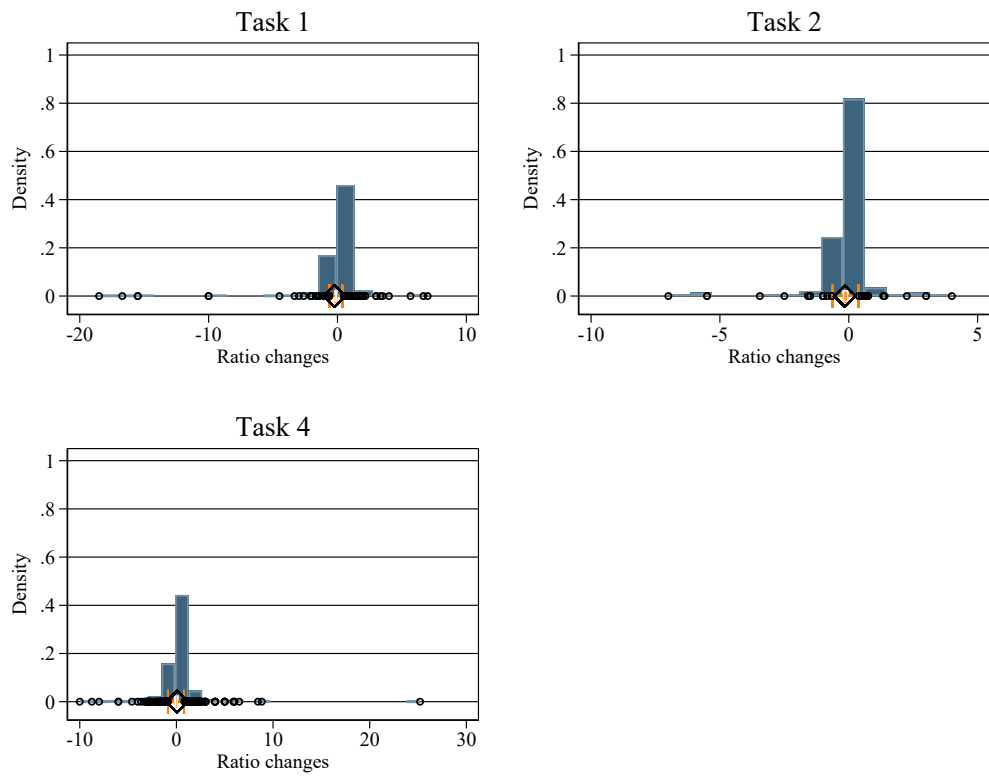


Figure A6: Residuals and estimates of a new color before and after observing a new color, second color.

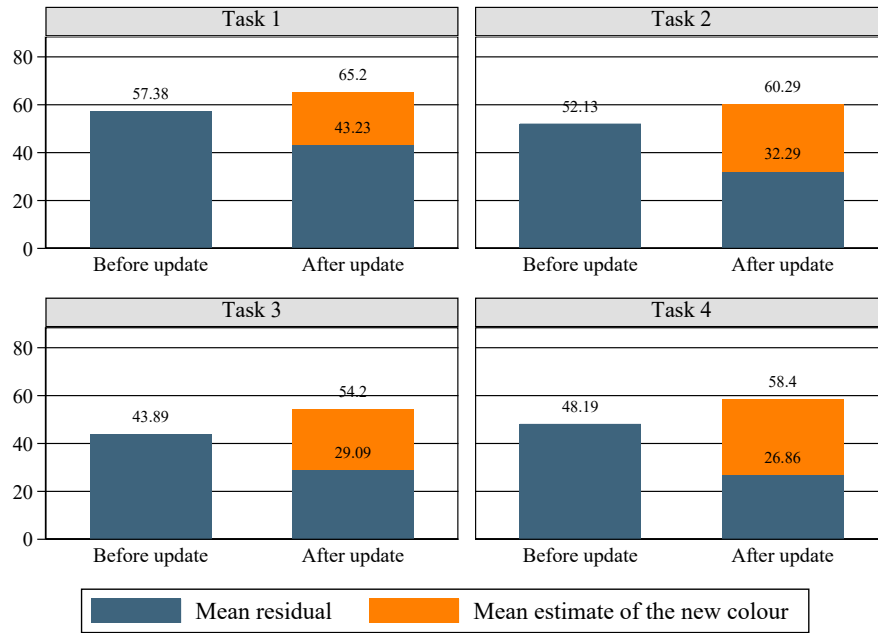


Figure A7: Residuals and estimates of a new color before and after observing a new color, third color.

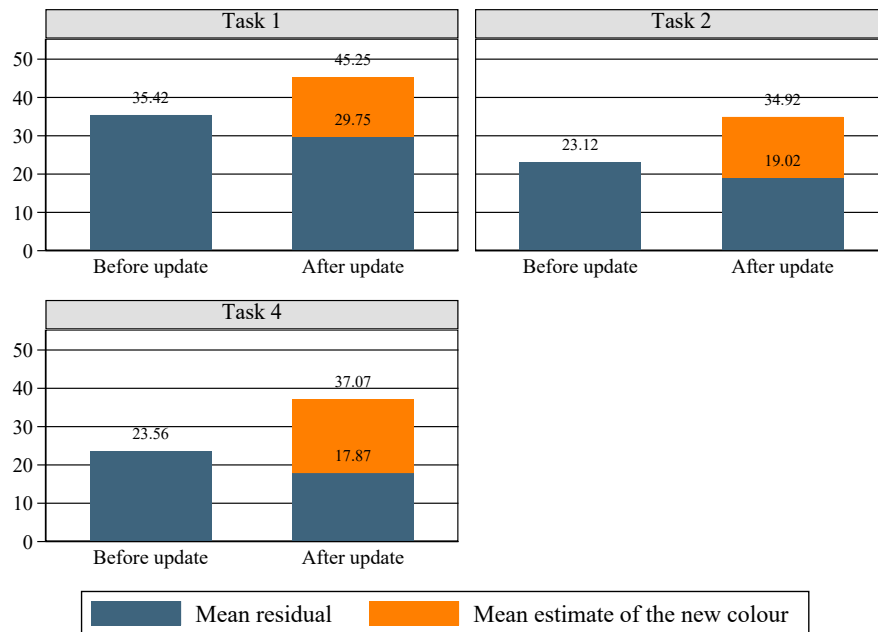


Table A5: Differences in the ratio changes, depending on whether participants expected a surprise ($\hat{p}_x^t > 0$).

		Didn't expect surprise	Expected surprise	p-value	p-value (corr)
Task 1	ΔR^3	5	80	0.664	1.000
	ΔR_1^4	5	79	0.919	1.000
	ΔR_2^4	5	79	0.202	1.000
	ΔR_3^4	5	79	0.466	1.000
Pooled	ΔR_P^4	15	237	0.221	1.000
Task 2	ΔR^3	54	115	0.698	1.000
Task 4	ΔR^3	38	135	0.096	1.000
	ΔR_1^4	33	131	0.330	1.000
	ΔR_2^4	33	130	0.430	1.000
	ΔR_3^4	33	130	0.343	1.000
Pooled	ΔR_P^4	99	391	0.134	1.000

Wilcoxon rank-sum test, p-values corrected by Bonferroni-Holm procedure.

Table A6: Regression: relation between the number of sampled marbles and the residuals, panel fixed effects.

Ratio differences ΔR	
Num. draws	-0.735** (0.046)
Num. colours observed	-8.495** (0.532)
Constant	52.451** (1.868)
Observations	20,880
Subjects	174

* $p < 0.05$; ** $p < 0.01$

Standard errors in parentheses, standard errors are clustered at the individual level.

Table A7: Regression: relation between the number of sampled marbles and the residuals after sample round 15, panel GLS.

Ratio differences ΔR	
Num. draws	-0.349** (0.051)
Num. colours observed	0.852 (0.795)
Cognitive ability	-2.299** (0.517)
Four colours first	-2.800 (2.250)
Constant	35.562** (11.238)
Observations	9,900
Subjects	165

* $p < 0.05$; ** $p < 0.01$

Standard errors in parentheses, standard errors are clustered at the individual level.

Note: The estimation additionally controls for age, gender, being an economics student and risk aversion but the coefficients are not reported.